LIBRARY OF PRINCETON
7. 21. 49.
THEOLOGICAL SEMINARY

HM251
.T45
v. 1

LIBRARY OF PRINCETON
JUL 21 1949
THEOLOGICAL SEMINARY

OBSERVATIONAL STUDIES OF SOCIAL BEHAVIOR

VOLUME I—SOCIAL BEHAVIOR PATTERNS

BY

DOROTHY SWAINE THOMAS

ALICE M. LOOMIS

RUTH E. ARRINGTON

with the assistance of

ELEANOR C. ISBELL

INSTITUTE OF HUMAN RELATIONS
YALE UNIVERSITY
1933

PRINTED IN THE UNITED STATES OF AMERICA

FOREWORD

This volume is a report of a particular stage in a specific program. It is our own working copy, presented in this form at this time, partly because the growing interest in the possibilities of observational techniques leads us to believe that our results, however tentative, may be useful to other workers, and partly in the hope that this rather detailed presentation may bring helpful criticism. The report is in the nature of a case study of the point to point development and present status of our own program. The interdependence of our techniques and findings with certain of the studies in related behavior fields, and some of their possible implications for these fields, will be analyzed and discussed in a subsequent volume.

It will be obvious to the reader that we have become involved in many difficulties in our attempt to develop precise measurements of social interaction. It will also be clear that we have by no means reached a satisfactory solution of many of them.

We are indebted to many persons for advice and coöperation. Professor Mark May, Professor E. B. Wilson and Miss Margaret Hogg have given invaluable advice, the results of which will, we hope, be apparent in our later analyses and reports. We have received especial favors from Professor H. M. Halverson of Yale University. Our actual observations were made possible through the coöperation of a New Haven industrialist, the foremen and workers in a factory, a member of the Board of Education, teachers in the public schools, the Y. W. C. A. and the Paramount-Publix Corporation. The studies described in Chapter I were made at the Child Development Institute, Columbia University.

Digitized by the Internet Archive
in 2025 with funding from
Princeton Theological Seminary Library

https://archive.org/details/observationalstu00thom

CONTENTS

TABLES

PART I

PART II

APPENDICES

CHARTS

PART I

PART II

APPENDICES

FIGURES

PART I

PART II

INTRODUCTION

Our general interest, in common with that of many other sociologists, is in the study of the patterns or the consistent modes of behavior of individuals; in the relation of the behavior patterns of individuals to those of other interacting individuals; in the degree to which groups of interacting individuals exhibit common patterns; in the extent of individual deviation from group patterns; and, finally, in the relationship between such group patterns and individual deviations on the one hand and between the group pattern and social structure on the other hand. These interests involve further the study of the variation of patterns within a given structure and the degree of change of patterns concomitant with changes in social structure.

Our efforts have been specifically directed toward obtaining some simple statistical indices of the kinds, amount, range and degree of variation of social activity of individuals in "normal" situations of differing degrees of structuralization (school, work and recreation groups, etc.).

We have chosen the method of direct observation of individuals and groups, and have attempted to derive differentiating indices of social behavior wholly in terms of our observational records of some of its overt aspects. We assume that such records may be indicative of the types of contacts individuals make with other persons and receive from them, the successes and failures of these contacts as reflected in the behavior of interacting individuals, the extent to which interpersonal activities dominate in certain individuals to the apparent exclusion of interest in the material environment, the extent to which other individuals show withdrawal tendencies from other persons in their immediate environment and similar aspects of behavior. We further assume that such observational records can be so controlled as to lend themselves to statistical manipulation and lead to a quantitative differentiation of individuals and groups.

It should be emphasized that we have chosen a very small part of the general field of sociological interest, that we have defined

our specific problems in terms of limited overt behavior patterns capable of direct observation and, further, that we are immediately concerned only with that part of our data which lends itself to statistical analysis.

Our evaluation of the indices thus derived depends partly on the adequacy of our definitions of relevant activities in terms of their simple overt components, and partly on our success in devising techniques which make the observer an efficient recording instrument.

Part I of this volume represents our program of observational studies, in which an attempt is made to develop simple indices of social interaction and to define individual and group behavior patterns in terms of these indices. The first chapter deals with the development of a technique in a situation which in and of itself imposed relatively slight control on the behavior of the individuals concerned. The patterns defined in this study, then, are to a very slight extent predetermined by the structure of the immediate situation. The following chapters show the development of similar techniques in situations which represent varying degrees of structuralization, i. e., where a predetermined job in and of itself defines certain of the patterns of behavior, and where an attempt must be made to separate those aspects of the behavior patterns which are determined by the structure of the immediate situation from those that are not so determined. The usefulness of observational techniques in producing significantly differentiating indices will, in the long run, undoubtedly not be the same in the various types of situations which we discuss in these chapters.

Part II represents a more rigid test of the capabilities and deficiencies of the observer's use of our techniques than it was possible to obtain in the situations we were studying. The results of observations of characters in moving picture films are described and analyzed to throw light upon the degree of reliability of our techniques, the gross errors of observation, the biased errors of individual observers, the influence upon reliability of the specific situation observed, and, finally, the deficiencies of the records of ordinarily observed behavior as compared with records more nearly approximating "true" behavior.

Volume II, which is already partly prepared, will deal with

further data on reliability and will consider particularly the differential factor of an observer's self consistency. It will also present the results obtained by a method of recording where one of the variables (the timing error) has been more completely controlled through mechanical aids.

PART I

TECHNIQUES AND RESULTS

CHAPTER I

SOCIAL-MATERIAL-SELF ACTIVITY PATTERNS IN A NURSERY SCHOOL GROUP

I

THE modern nursery school represents an ideal situation in which to study the social behavior of young children. Here is an environment rich in material equipment where the young child is thrown into the society of his peers in such a way as to produce competition for the enjoyment of materials and coöperation in their use. Certain routine activities occur, *e. g.*, rest periods, the lunch period, story and music periods, but the greater part of every morning is given over to a so-called "free play" period during which the child's activities are relatively undirected by adults and his choices can develop into patterns determined largely by his interrelationships with those of his own age level. He may successfully corner certain materials and direct his social efforts to maintaining them for his own use; he may become the focus of the social advances of others and respond in kind, or he may withdraw to an isolation apparently satisfying in itself and independent of the materials and persons in the environment. To what extent do the undirected activities of a child fall into patterns indicative of his social-material-self interests? To what extent are the individual patterns common to the group? To what extent do personality deviates emerge in relation to group patterns?

To answer these questions, the younger group in the nursery school of the Child Development Institute of Teachers College, Columbia University, was studied during the months February to June, 1930. This group was composed of twenty children, ten boys and ten girls, ranging in age (as of February) from sixteen to thirty-two months. A minimum of twenty-four five-minute records was taken on each child during this period with a sampling procedure designed to insure a representative picture of his total activity. The complete results of this study are

published in Child Development Monograph, No. 8.[1] It is necessary to summarize the technique and the results here, both in order to show what types of behavior patterns emerged at this age level as compared with those shown in our later studies at the kindergarten level, and to show the fundamentals of the technique which we have used, with appropriate modifications, in all of our studies.

The technique used is described as follows and illustrated in the accompanying chart and activity code.

Mimeographed record blanks were divided into five-second intervals, corresponding to intervals on the stop-watch. Watches were started at the beginning of the observation and allowed to run continuously until the end of the five-minute period. The record blank itself provided an automatic check on accuracy of timing. It was [not] necessary to record at the time of observation the number of seconds spent in a given activity; it was necessary only to indicate by lines and symbols the duration or frequency of activities within specific time intervals. This device reduced the mechanical process of recording to an extremely simple form and made [quantification possible]. . .

[The child's activities were divided] into two parts, to be recorded simultaneously by two observers. . . . It was decided that one observer could best record the three mutually exclusive types of activity which it had been possible to time in terms of duration, namely, contact with material, physical activity, and no overt activity; while the other recorded talking, physical contacts, laughing, and crying in terms of [occurrence] within intervals of five seconds.

Continuous lines were drawn in the appropriate column to indicate duration of the activity to the nearest second, and dotted lines in the "material" column to indicate inactive contact with material (lying on the slide, sitting in the wagon, etc.). Each line was labeled with the name of the material used, and a separate line was used for each material. Names of materials merely touched in passing were underlined and arbitrarily assigned a time value of a second. Symbols were used to represent the initiator or recipient of physical contacts and to distinguish between talking to persons, talking to self, and vocalizing. The following key will serve to explain the accompanying sample record of Child B. [Chart I]

M—Material
P—Physical activity
NO—No overt activity
T—Talking

[1] Arrington, Ruth E., *Interrelations in the Behavior of Young Children*, Child Development Monographs, No. 8. Bureau of Publications, Teachers College, Columbia University. 1932.

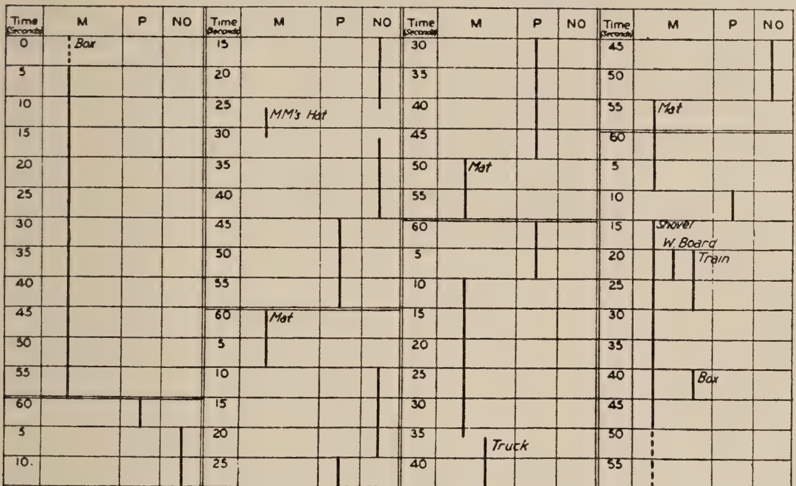
Chart I. Sample Five-Minute Record of the Undirected Activities of Child B

FORM A

MATERIAL-SOCIAL-SELF ACTIVITIES RECORD

CHILD *B* TIME *9⁴² – 9⁴⁷* DATE *2/25/30* RECORDER *MSF*

CHILDREN PRESENT: A-AH-B-BD-D-E-EB-G-H-I-J-M-MK-MM-N-P-PH-R-RS-V



FORM B

MATERIAL-SOCIAL-SELF ACTIVITIES RECORD

CHILD *B* TIME *9⁴² – 9⁴⁷* DATE *2/25/30* RECORDER *REA*

CHILDREN PRESENT: A-AH-B-BD-D-E-EB-G-H-I-J-M-MK-MM-N-P-PH-R-RB-V

Time (Seconds)	T	PC	L	C	Time (Seconds)	T	PC	L	C	Time (Seconds)	T	PC	L	C	Time (Seconds)	T	PC	L	C
0			✓		15		EB			30			✓		45				
5	✓		✓		20					35			✓		50	✓			
10			✓		25		MM			40			✓		55	Δ			
15					30					45					60				
20	✓		✓		35					50			✓		5				
25					40		EB			55	✓		✓		10	✓			
30					45		EB			60	✓				15				✓
35	✓		✓		50					5	Δ				20				
40		H			55					10			✓		25	✓			
45	✓				60		EB			15	✓		✓		30				
50					5					20					35	✓			
55			✓		10					25					40	✓			
60					15		BD	✓		30			✓		45		N		
5		H			20			✓		35	✓		✓		50				
10		EB			25		BD			40	✓				55				

PC—Physical contact
L—Laughing
C—Crying
Δ—Teacher
π—Group
∨—Occurrence of talking (not addressed to a particular person or group of persons), laughing, or crying
∧—Vocalizing
O—Observed child the object of physical contact (*e. g.*,—R—contact with observed child initiated by R)
(Talking to persons was indicated by the initials of the child, or the symbol for the teacher or group).

Thus, interpreting Child B's record for the five-minute period beginning at 9:42 A. M. on February 25, as shown in Chart [I], Forms A and B, we see that nineteen children were present on the roof when the record was begun. Child B was physically inactive, though in contact with material, in the first interval of the first minute, and again during the last ten seconds of the fifth minute, making a total of 15″. He used material 160″, spent 60″ in physical activity, exclusive of the time involved in using material, and 60″ in no overt bodily activity. He talked to himself . . . during fourteen time intervals, addressed a teacher twice, initiated physical contacts with Children H, EB, MM, BD, and N and received contacts from Children H, EB, and BD, and laughed during eighteen intervals.[1]

The specific definitions of the behavior to be included in each category follow.

Use of material was defined as active manipulation with hands, feet, or other parts of the body. Five general types of material were included in the definition: (1) play equipment; (2) additional nursery school equipment used in attending to the child's physical needs, such as paper towels, cups, etc.; (3) clothing, when actively manipulated; (4) furniture (chairs, tables, drinking fountain, etc.); and (5) parts of the building (wall, posts, iron grating, doors, window ledges, etc.). No attempt was made to differentiate degrees of activity in relation to specific materials. A child who stood beside the swing, gently moving the rope back and forth, was considered to be using the swing, as well as the child who actively swung himself. . . .

Physical activity included walking, running, jumping, rolling, crawling, moving up and down in position, waving arms, clapping hands, falling down, and active physical contacts that did not involve material. Obviously, this category does not measure the child's total physical activity,

[1] Arrington, *op. cit.*, pp. 22–25.

since activity directed toward material is not included. It is merely a gauge of the proportion of activity not concerned with material objects, which, combined with the proportion of time spent in active use of material, affords a gross measure of total time spent in bodily activity as contrasted with time spent in inactivity. A more exact distinction between the highly active and the slightly active child will have to be based upon a special study of physical activity as a [category] in itself.

The third category, which has been called no overt activity, was defined as standing, sitting, or lying still, making no observable reaction to the environment other than looking around, or looking at persons or things. No attempt was made to decide whether the child was day-dreaming, or whether his attention was directed toward persons or things. Definitely directed activity, in the relatively few instances in which it occurred, was recorded as no overt activity on one blank, and as an object physical contact on the other. Thus, when a teacher attended to the child's needs, such as tying shoes, wiping nose, etc., or, in rare cases, carrying the child from one place to another, and the child made no observable response to the physical contact, the lack of response was indicated under the heading "no overt activity."[1]

Under physical contacts, in addition to the obvious contacts of pushing, hitting, embracing, fighting, etc., were grouped things done for the child by the teacher, such as nose wiping, fixing of clothes, etc., being used as means of support or vice versa, and certain contacts involving material that should logically have been placed as a sub-class under "use of material with persons," namely, taking things away from other children, or having things taken away, giving things to others, or receiving things from others. An incident was defined as a physical contact through material only when the material involved was of such a nature as to be handed from child to child, or from child to teacher. One exception to this definition was made in the case of the swing. Actual physical contact with the child in the swing or with the child who tried to take the swing away seemed so inevitable that this activity was included, although taking wagons, doll carriages, trucks, etc., was not unless an actual body contact occurred.[2]

Contacts lasting more than five seconds were indicated by a continuous line below the symbol for the person with whom the contact was made. . .

Talking to persons was defined as using the person's name in direct address, or talking while using the same material as other children, or different materials, in close proximity. If the observer was uncertain as to whether or not speech was directed toward a particular person in a group, the remark was assumed to be addressed to the group. As is

[1] Arrington, *op. cit.*, pp. 25–26.
[2] Arrington, *op. cit.*, pp. 16–17.

frequently the case with an indefinite category of this sort, it is probable that an overweighting of talking to the group resulted from the difficulty of determining whether the child was talking to himself or to other persons, and of determining to whom his speech was addressed. Vocalizing was interpreted as making audible sounds that had no meaning for the observer.[1]

The reason for the inclusion and strict delimitation of each category and the exclusion of other aspects of behavior will become clearer in the second half of this chapter when we discuss the derivation of the technique and the validation of these categories in terms of independent studies. This technique provided a simple and clear-cut basis for a differentiation of children in terms of their functional preoccupation with the material and social resources of their environment. It showed that, for the group,

on an average, 245 seconds out of every 300 seconds [were spent] in contact with material, 209 seconds in active use of material, 36 seconds in inactive contact with material, 37 seconds in physical activity not involving material, and 18 seconds in no overt activity. In other words, 82% of the time . . . was spent in contact with material, 70% in active use of material, 12% in inactive contact with material, 12% in physical activity, and 6% in no overt activity.[2]

The actual number of time units of social and self activities (as defined by our units of talking to persons, physical contacts with persons, talking to self, laughing and crying) was capable of computation. For the group as a whole

the infrequency of the social and self activities, as narrowly defined in the study, is worthy of note. The average occurrence, for the group, of talking to persons, was at the rate of 3.2 five-second intervals in 60; that is, these children tended to talk to other children or to adults [slightly] more than 5% of the time they were observed. They were in physical contact (self-initiated) with persons even less frequently, at the rate of 2.2 intervals in 60, or 4% of the time observed. They were recipients of physical contacts from other persons in 3.4 intervals in 60, or 6% of the time. The average ratio of subject to object contacts (with children), for the group, was 1.28. Of the self activities, talking, not specifically addressed to persons, occurred with greatest frequency, at the rate of 4.7 intervals in 60, or 8% of the time. Laughing occurred at the rate of 1

[1] Arrington, *op. cit.*, pp. 26–27.

[2] Arrington, *op. cit.*, pp. 83–84.

interval in 60, or 2% of the time observed, and crying, at the rate of 1 interval in 120, or 1% of the time.[1]

With "social contacts" defined in terms of the physical and the verbal, the technique enabled us to relate the social and self activities to the concomitant basic activities of material manipulation, physical activity and no overt activity. Seventy-six per cent of the time of this group was spent in manipulating material without any social contact; 10.7%, in contact with materials accompanied by social contact; 3%, in "purely social" contact (accompanied, that is, by a physical activity or no overt activity record) and 10.7%, in activities involving no observable contact either with persons or with things. A check by the technique on another nursery school group of the same age range in the same school for another year showed no significant differences in the group means, indicating the possibility of applying such a technique to the derivation of behavior norms in this field.

We are interested also in the deviates from the central tendency of the group, and the range of individual difference.

Significant deviations from the group averages were noted for individual children in all the types of behavior investigated. The range for total time spent in use of material was from 50% to 88%; for physical activity not involving material, from 6% to 28%; for no overt activity, in contact with material, from 4% to 21%; and for no overt activity without contact with material, from 2% to 17%. Priscilla, for example, spent 92% of the time she was observed in contact with material and 88% in active use of material. Albert, at the opposite extreme, spent only 50% of his time in using material and 11% in inactive contact with things. The range for gross physical activity was from 68% to 94%. Priscilla was inactive only 6% of the time observed, whereas Rebecca spent 32% of her time passively watching the activities of others.

While talking occurred on an average in about 13% of the time intervals, for the group, one child talked during only 1.8 intervals out of 60, or 3% of the time; another, in 21.7 intervals in 60, or 36% of the time. For frequency of talking to self, the range was from 1% to 28%; for talking to persons, from 1% to 8%. For initiated physical contacts, there was a range from 1% to 6%; and for contacts received, from 3% to 8%. While the average ratio of subject to object physical contacts was . . . greater than one (1.28), Pauline made only about a third as many contacts with children as she received from them, and Margery initiated

[1] Arrington, *op. cit.*, pp. 86–87.

contacts with children almost two and a half times as often as she received them.

Extreme differences are also to be noted in the types of social contact initiated most frequently by individual children, and in the recipients of the contacts made by different children. Of Priscilla's social contacts, 88% were of the verbal type. Of Amy's contacts, 76% were physical. Of Pauline's contacts, 85% were directed toward adults. Amy, Horace, Grace, and Rudolph made contacts predominantly with other children.

Although laughing occurred on an average during only 2% of the time the group was observed, Benny laughed during 13% of the time intervals in which he was observed. His extreme deviation from the group norm is shown by the fact that the child who ranked second in frequency of laughter laughed during only 4% of the intervals in which he was observed. For two children, Albert and Basil, no laughter was noted during the entire observation period.

Crying occurred with extreme infrequency in the group as a whole. The child who ranked highest in frequency of crying cried during only 3% of the total time observed. It is significant that this child, Emery, also ranked second highest in frequency of laughter. This suggests a behavior pattern of high "emotionality," or emotional instability. Benny, on the other hand, for whom most instances of laughter were recorded, had no records of crying at all. In his case, the extreme deviation from the norm for the group with respect to this type of behavior seemed to indicate a behavior pattern adopted as a means of attracting attention from the group rather than to be a symptom of emotional instability.[1]

A technique of this sort also makes possible an accurate study of change in the behavior of the same individuals and groups over a period of time, the so-called longitudinal or genetic approach. Lack of data up to the present has rendered impossible any but a crude test along these lines. Sixteen of the children studied by this technique were studied again the following year, and their relative positions in the group were compared. As might be expected, consistency (and its lack) is itself a highly significant behavior trait which is manifested to varying degrees in the records of individual children. For the group as a whole the greatest consistency (as measured by rank correlations) occurred in talking and laughing, the next greatest in physical contacts,—a slight indication of a general tendency to individual

[1] Arrington, *op. cit.*, pp. 90–92.

stability in these social manifestations. The other measures of behavior showed little group consistency.

The definition of personality deviates with the accumulation of data of this sort is possible.

We might expect to find in any unselected group of individuals some whose interests are evenly divided between people and things, who spend much time in active manipulation of material, but who also initiate frequent social contacts with their fellows. These we would probably consider the best "adjusted," since adult life, in society as at present organized, demands the ability to control both the material and the social factors in the environment. We would expect to find others for whom the manipulation of things, the mastery of the inanimate factors in the environment, is of more absorbing interest than social intercourse. Still others would be found whose concern with mere things is largely casual, but whose dominant interest is in association with other human beings, in watching, participating in, and controlling their activities. A fourth group of individuals, conspicuous for their lack of overtly expressed interest either in things or people, we are wont to diagnose as abnormal, on the assumption, unproved though it is, that normal behavior involves an even division between the material and the social. . . . We may use the term "material-social" for the first group, "material-non-social" for the second, "non-material-social" for the third, and "non-material-non-social" for the extreme deviates, interpreting these labels to mean relatively high or low frequency, rather than presence or absence of material and social activity.[1]

The technique gives a basis for such a differentiation on the nursery school level, if we can show that the elements we have selected are representative, to any marked degree, of these more general behavior manifestations.

Because of the limited data at this level we have made no attempt to develop this point further statistically. The accompanying samples of actual records on four children indicate possible extremes of behavior patterns in these four "types,"—Grace, the material-social, simultaneously engaged in active manipulation of materials and in extensive verbal and physical contacts with other children (see p. 2, *supra*, for activity code); Eunice, the material non-social, as active in material manipulation as Grace, never addressing a remark to her peers or making contacts with them, but accompanying her material activity

[1] Arrington, *op. cit.*, pp. 133–134.

CHART II. SAMPLE FIVE-MINUTE RECORD OF THE UNDIRECTED ACTIVITIES OF CHILD GRACE

FORM A

MATERIAL - SOCIAL - SELF ACTIVITIES RECORD

CHILD GRACE TIME 10^{10} – 10^{15} DATE 2/16/31 RECORDER LBC

CHILDREN PRESENT: A - AH - BD - D - DS - E - EB - ED - G - H - HH - I - J - M - MK - MM - RB - S - V - VA

Time (Seconds)	M	P	NO	Time (Seconds)	M	P	NO	Time (Seconds)	M	P	NO	Time (Seconds)	M	P	NO
0	Cage			15	Box			30	Box			45	Box		
5				20				35				50			
10				25				40				55			
15				30				45				60			
20				35				50				5			
25				40				55				10			
30				45				60				15			
35				50				5				20			
40				55				10				25			
45				60				15				30			
50	Box			5				20				35			
55				10				25				40			
60				15				30				45			
5				20				35				50			
10				25				40				55			

FORM B

MATERIAL - SOCIAL - SELF ACTIVITIES RECORD

CHILD GRACE TIME 10^{10} – 10^{15} DATE 2/16/31 RECORDER MEG

CHILDREN PRESENT: A - AH - BD - D - DS - E - EB - ED - G - H - HH - I - J - M - MK - MM - RB - S - V - VA

Time (Seconds)	T	PC	L	C	Time (Seconds)	T	PC	L	C	Time (Seconds)	T	PC	L	C	Time (Seconds)	T	PC	L	C
0					15					30	TT		✓		45				
5					20	MM				35	TT				50	MM			
10					25	MM				40	TT				55	MM			
15					30					45		MM			60	EB	EB		
20					35					50					5	EB			
25	TT				40	TT				55					10	EB			
30	MM	MM			45					60					15	EB			
35	MM				50	MM	MM			5					20	MK			
40					55	MM				10					25	MK			
45					60	MM				15	MM	MK			30	TT			
50	MM	MM			5	MM				20	MM				35	MK	MK		
55	MK				10	MM				25	TT				40	MK			
60	MK				15					30	TT				45	MK			
5	MM	MM MM			20	MK				35					50	TT			
10	MM	MM			25					40	MK	MK			55	TT			

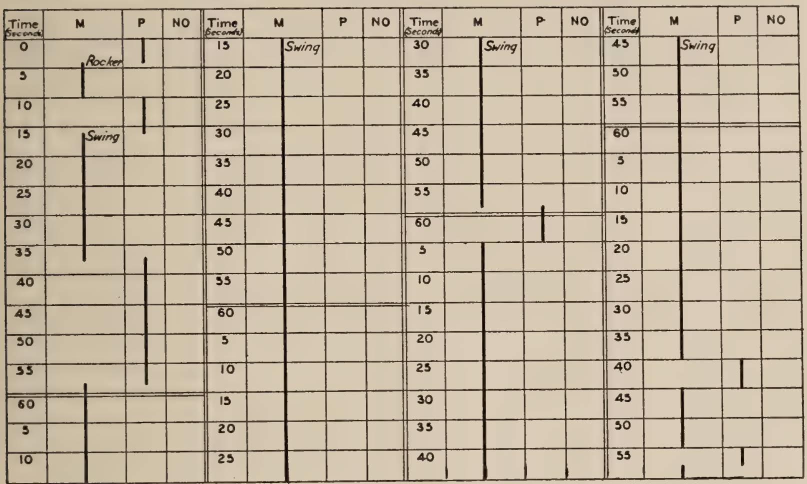
Chart III. Sample Five-Minute Record of the Undirected Activities of Child Eunice

FORM A

MATERIAL - SOCIAL - SELF ACTIVITIES RECORD

CHILD *EUNICE* TIME *10³¹ - 10³⁶* DATE *3/19/31* RECORDER *LBC*

CHILDREN PRESENT: A - AH - BD - D - DS - E - EB - ED - G - H - HH - I - J M - MK - MM - RB - S - V - VA



FORM B

MATERIAL - SOCIAL - SELF ACTIVITIES RECORD

CHILD *EUNICE* TIME *10³¹ - 10³⁶* DATE *3/19/31* RECORDER *MEG*

CHILDREN PRESENT: A - AH - BD - D - DS - E EB - ED - G - H - HH - I - J - M - MK - MM - RB - S - V - VA

Time (Seconds)	T	PC	L	C	Time (Seconds)	T	PC	L	C	Time (Seconds)	T	PC	L	C	Time (Seconds)	T	PC	L	C
0	✓				15					30	✓				45	△			
5	✓				20					35	✓				50	△			
10	✓				25	✓				40	✓				55	△			
15					30	✓				45	✓				60	△			
20	✓				35					50	✓				5	△			
25					40	✓				55	△				10	△			
30	✓				45	✓				60	△				15	△			
35	✓				50	✓				5					20	△			
40	✓				55					10	△				25	△			
45					60	✓				15					30	△			
50					5	✓				20	△				35				
55	✓				10	✓				25	△				40	△			
60	✓				15	✓				30	△				45				
5	✓				20					35	△				50			✓	
10	✓				25	✓				40					55	△			

CHART IV. SAMPLE FIVE-MINUTE RECORD OF THE UNDIRECTED ACTIVITIES OF CHILD REBECCA

FORM A

MATERIAL-SOCIAL-SELF ACTIVITIES RECORD

CHILD *REBECCA* TIME *9[35] – 9[40]* DATE *4/20/31* RECORDER *LBC*

CHILDREN PRESENT: A - AH - BD - D - DS - E - EB - ED - G - H - HH - I - J - M - MK - MM - RB - S - V - VA

Time (Seconds)	M	P	NO	Time (Seconds)	M	P	NO	Time (Seconds)	M	P	NO	Time (Seconds)	M	P	NO
0				15				30				45			
5	*Board*			20				35				50			
10				25				40				55			
15				30				45				60			
20				35	*Medal*			50				5			
25				40				55				10			
30				45				60				15			
35				50				5				20			
40				55				10				25			
45	*Card*			60	*Sand*			15				30			
50				5				20				35			
55				10				25				40			
60				15				30				45			
5				20	*Railing*			35				50			
10				25				40				55			

FORM B

MATERIAL-SOCIAL-SELF ACTIVITIES RECORD

CHILD *REBECCA* TIME *9[35] – 9[40]* DATE *4/20/31* RECORDER *MEG*

CHILDREN PRESENT: A-AH-BD-D-DS-E-EB-ED-G-H-HH-I-J-M-MK-MM-RB-S-V-VA V=Visitor

Time (Seconds)	T	PC	L	C	Time (Seconds)	T	PC	L	C	Time (Seconds)	T	PC	L	C	Time (Seconds)	T	PC	L	C
0					15					30	V				45				
5					20		*VA*			35					50				
10					25	*VA*				40					55	*MM*	*MM*		
15	V				30					45	V				60	*MM*			
20	V				35	*VA*	*VA VA*			50					5				
25					40					55					10		*H H*		
30					45					60					15				
35					50					5					20				
40					55					10	Δ				25				
45	*EB*				60					15					30	*ED*			
50	*EB*				5					20					35	*ED*			
55		*EB EB*			10		*HH*			25		Δ			40				
60					15	✓	*HH*			30					45	*ED*			
5					20					35					50				
10					25	V				40					55	*ED*	*ED ED*		

CHART V. SAMPLE FIVE-MINUTE RECORD OF THE UNDIRECTED ACTIVITIES OF CHILD ALBERT

FORM A

MATERIAL-SOCIAL-SELF ACTIVITIES RECORD

CHILD ALBERT TIME 10^{25} 10^{30} DATE 2/20/30 RECORDER M/M

CHILDREN PRESENT: A-AH-B-BD-D-E-EB-G-H-I-J-M-MK-MM-N-P-PH-R-RB-V

Time (Seconds)	M	P	NO	Time (Seconds)	M	P	NO	Time (Seconds)	M	P	NO	Time (Seconds)	M	P	NO
0				15				30	Jungle Gym			45	Jungle Gym		
5				20				35				50			
10				25				40				55			
15				30				45				60			
20				35				50				5			
25				40				55				10			
30				45				60				15			
35				50				5				20			
40				55				10				25			
45				60	Jungle Gym			15				30			
50				5				20				35			
55				10				25				40			
60				15				30				45			
5				20				35				50			
10				25				40				55			

FORM B

MATERIAL-SOCIAL-SELF ACTIVITIES RECORD

CHILD ALBERT TIME 10^{25}– 10^{30} DATE 2/20/30 RECORDER REA

CHILDREN PRESENT: A-AH-B-BD-D-E-EB-G-H-I-J-M-MK-MM-N-P-PH-R-RB-V

Time (Seconds)	T	PC	L	C	Time (Seconds)	T	PC	L	C	Time (Seconds)	T	PC	L	C	Time (Seconds)	T	PC	L	C
0					15					30					45		△		
5					20					35					50				
10					25					40	✓	△			55				
15					30	✓				45					60				
20					35					50					5		△		
25					40					55		△			10				
30					45	✓				60					15		△		
35	✓				50					5					20				
40					55		△			10					25				
45					60					15	△				30				
50					5					20					35				
55					10					25					40				
60	✓				15					30	△				45				
5	✓				20					35					50				
10					25		△			40					55	△			

with a tremendous flow of monologue and of remarks addressed to the teacher; Rebecca, the non-material social, manipulating materials casually and non-continuously, but talking and making and receiving physical contacts with no less than five children and two adults in a five-minute period; Albert, the non-material-non-social, wandering around and talking to himself when left to his own devices and manipulating materials only when actually held in contact with them by the teacher.

This attempt to define basic behavior patterns showing the child's involvement with things, with persons and with himself can best be clarified in terms of certain earlier attempts to define such activity patterns by other techniques.

II

A comparison of the earliest technique[1] in this program (Barker's) with the one just described shows the progressive improvement in control as our definitions became clarified and indicates the processes involved in deriving a workable technique.

Barker's method differed from the uncontrolled diary record in three respects. (1) It involved a graphic record of the movements of the child under observation. Toward *what* material objects located *where* did the child direct his movements? In *what* groups composed of *what* people did he involve himself, and *what* people became involved with him? (2) It limited the diary notations to descriptions of the shifts in the child's activity as symbolized on the chart of his movements. (3) It timed the duration of the successive acts. The observer traced the child's progress on a floor plan for five minutes, indicated at what points in this progress materials were manipulated and personal contacts were made, and timed the duration of each discrete activity. An actual five-minute record and the activity code are reproduced on page 15.

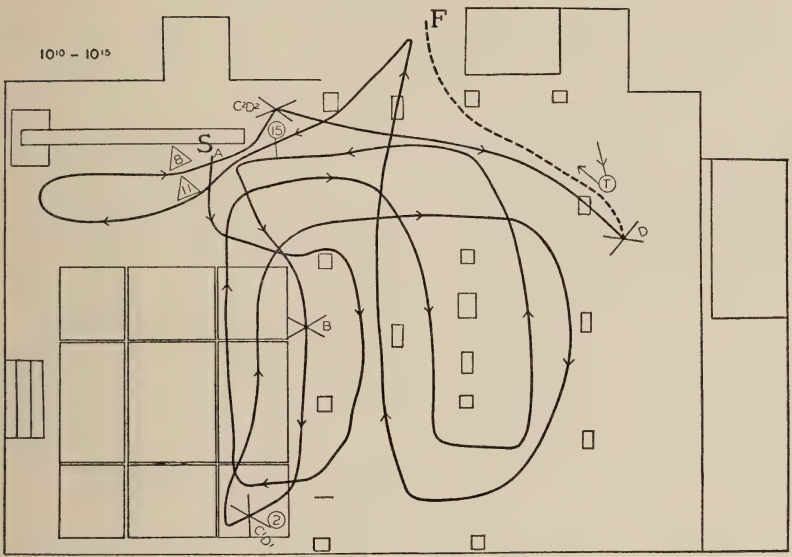
Although many five-minute records were taken on each of the group of children very little information pertinent to their social behavior was obtained because of two technical difficulties. (1) The reliability of the technique, as described later, was so

[1] Barker, Margaret, *A Technique for Studying Social-Material Activities of Young Children*, Child Development Monographs, No. 3. Bureau of Publications, Teachers College, Columbia University, 1930.

low on all points where it had attempted to differentiate individuals on the basis of their social contacts that any generalizations were invalid. (2) The *a priori* delimitation of classes of activities into material, social and self components had not been

CHART VI. SAMPLE FIVE-MINUTE RECORD OF THE UNDIRECTED ACTIVITIES OF CHILD EDWARD

From Barker, *A Technique for Studying the Social-Material Activities of Young Children*, p. 8, Child Development Monographs, No. 3.



Edward—12/6/28—M.B.

AB—Trying to take cover from Alma to give to Ruth. 35″

CD—One foot on tricycle, pushing it with other. Then Henry comes up and sits on it and Henry pushes him. Falls off but doesn't cry. Gets on again. 243″

C¹D¹—Katherine says, "You knocked me down." He looks. 7″

C²D²—Looks at Ruth, who is crying. Goes downstairs at end of five-minute period. 10″

adequately carried through, with the result that most of the activities were recorded as comprising at least two of these three components. With regard to the latter point, Barker attempted to classify the diary notations *a posteriori* according to a schematization which Arrington had evolved in terms such that the social elements in each activity would be thrown into focus.

All activities [were divided] into two main groups: those having to do with material, and those in which no material was involved. Each group had two subdivisions, social and non-social. All activities were thus separated into four mutually exclusive groups: (1) the *material-social* activities, including those concerned simultaneously with things and with persons; (2) the *material* activities, including those concerned with material alone; (3) the *social* activities, including those concerned predominantly with persons; and (4) the so-called *self* activities, including those in which neither persons nor things were involved.[1]

The specific classification follows:

I. Material
- A. Social
  1. Observation of material with others, and of persons using material
     - Looking around, over roof, at things, with someone
     - Listening to story, music, etc.
     - Looking out into hall
  2. Approach to material
     - Going to sand box, jungle gym, etc., with someone
     - Going to sand box, jungle gym, etc., where children are playing
  3. Contact with material (persons involved spatially or functionally)
     - a. Accidental contact
       - Bumping into child on tricycle, etc.
     - b. Use of material with others, or in spatial contact with others.
- B. Non-Social
  1. Observation of material alone
     - Looking around, over roof, at things, alone
  2. Approach to material alone
  3. Contact with material
     - a. Accidental—bumping into wagon, falling down, etc.
     - b. Use of material alone

II. Non-Material
- A. Social
  1. Observation of persons
  2. Approach to persons, following persons, etc
  3. Verbal contact with persons

[1] Arrington, *op. cit.*, p. 15.

a. Initiated by child
b. Initiated by others
4. Physical contact with persons
a. Initiated by child
b. Initiated by others
5. Spatial contact with persons
Laughing, shouting with others, running around with others, standing at elevator with others, being approached by others

B. Non-Social
1. Daydreaming, sucking finger, etc.
2. Talking to self, singing by self, etc.
3. Crying, screaming
4. Kicking, crawling, rolling on floor, by self[1]

This classification was derived after a large number of records had been examined and the diary notations thrown into classes according to their apparent functional similarity. The crude empirical classes were:

I. Looking around
Sitting in wagon, etc.
Sitting on chair, etc.
Lying on back, etc.
Standing by person or thing
Daydreaming
Sucking finger

II. Sand box [activity]

III. Manipulation of small toys and other objects
Rocks in chair, pushes doll carriage, spins wheels of tricycle

IV. Talking to persons
Showing things to persons
Smiling at persons
Laughing at persons
Directing persons

V. (a) Looking at persons
Going up to persons
(b) Looking at things
Going up to things
Listening to story, music, etc.
Looking in cupboard, etc.

[1] Arrington, *op. cit.*, pp. 14–15.

VI. Activity [involving] big muscles
  Jungle gym, walking board, swing, slide, carrying blocks, kiddie kar, pulling big wagon, climbing steps
VII. Hammering
VIII. Activity in closets and boxes
IX. (a) Things done for the child by another person
  Nose wiped
  Pulled in wagon, etc.
  Directed by teacher or child
  Led by teacher or child
  Comforted by another
  Teacher intervenes in behalf of child
  . . .
  (b) Things are taken from child by [another] child
  (No defensive reaction)
  Teacher intervenes in behalf of another [child]
  (c) Things are given to child by another
X. Conflicts
  Aggressive or defensive contacts
XI. Accidents
XII. (a) Teasing actions
  (b) Friendly contacts
  Embraces someone, helps someone, laughs with someone, plays with someone, follows someone
XIII. Talks to self
  Sings
  Laughs by self
  Looks at self in mirror
XIV. Cries
  Screams
  Whimpers
XV. Attends to own wants
  Wipes nose
  Brushes sand off hands
  Arranges mittens, etc.[1]

Fortunately for our later progress, the impossibility of clarifying our problems by applying *a posteriori* classifications to our existing materials was soon demonstrated beyond a possibility of doubt.

In terms of the duration of activities, class I B, representing uncomplicated involvement with materials, accounted for 8.8%

[1] Barker, *op. cit.*, pp. 26–28.

of the total time of the group's activity, but was also found in combination with presumably "pure" social elements (I A or II A), 8.0% of the total time, with inseparable self elements (i. e., combined with II B) 1.5% of the total time, and in various inseparable combinations including both social and self elements 5.0% of the total time. Class II A, representing uncomplicated social interaction, accounted for 24.7% of the total time, but also appeared in combinations with inseparable material or self elements (or both) 24.9% of the time. Class II B, the uncomplicated self activities, accounted for only 0.7% of the total time of the group, but appeared in combination with material or social activities, in such a way that they could not be separated, 6.7% of the time. Class I A which was supposed to represent the simultaneous (logically and actually inseparable) material-social activities accounted for 33.3% of the total time, but also was combined with what were presumably separable elements no less than 25.8% of the total time. Thus the *a posteriori* classification led to a complete impasse in differentiating the group of children. The system itself would presumably have made a statistical differentiation possible, but the application of the classes to data previously collected proved impossible. Summarizing these results, the *a posteriori* classification of activities which had been recorded and timed by competent observers led to no statistical differentiation of the members of the group of children studied because (1) activities which were considered in this schematization to be genuinely inseparable occurred 33.3% of the time, (2) activities which were considered in this schematization to be genuinely separable were recorded in such a way as to be actually inseparable 32.6% of the time, and were separable only 34.2% of the time.

Even if the activities had been reliably timed (and this was strictly true only for the material activities), the technique could be considered as being informative about children's activities, but as leading to no possible useful indices which would differentiate their behavior patterns statistically.

The next step was, of course, so to define observational categories that a classification (according to *a priori* definitions) could be made at the time of observing. The graphic method was discarded as producing data irrelevant for our main purpose,

and attention was focused on timing and classifying the separable social-material-self behavior. The observer memorized a code and recorded the activities of the child symbolically as follows:

Code	*Description of Activity*
M S 1	Observation of material with persons
	Observation of persons and material
	Listening
M S 2	Approach to material with persons
	Approach to persons using material
M S 3	Accidental contact with material and persons
M S 4	Contact with material (persons involved)
M 1	Observation of material alone
M 2	Approach to material alone
M 3	Accidental contact with material
M 4	Contact with material alone
S 1	Observation of persons
S 2	Approach to persons
S 3	Verbal contact with persons
S 4	Physical contact with persons
S 5	Spatial contact with persons
N 1	Daydreaming
N 2	Talking to self, singing, vocalizing
N 3	Crying, screaming
N 4	Kicking, crawling, walking around, by self

Key: M S material-social; M material; S social; N non-material, non-social

Each activity was recorded in proper sequence, with the appropriate symbols indicating the main category to which it belonged (material, social, material-social, or self) and the type or degree of activity within the category. In each case the duration of the activity was recorded beside the symbols. For example, if the child was climbing on the jungle gym for a period of fifty seconds, the activity would be recorded as M 4 50″. If, during this activity, he talked to another child or had a physical contact with another person, the fact would be indicated as S 3 or S 4, plus the time and the symbols for the person or persons with whom the contacts were made.[1]

Very little experience with this method of recording was needed

[1] Arrington, *op. cit.*, pp. 18–19.

to show clearly that we had not yet overcome or even recognized one of the fundamental difficulties although our definitions had now been formulated clearly, so that a quantitative analysis of the data might have seemed justifiable. This fundamental difficulty was that the children could still not be adequately differentiated because of the clustering of a large part of their activity in the complex class, M S, (material-social) and the impossibility of assigning proportionate weights to the elements in this class. Furthermore, the technique gave unreliable results.

Arrington then made a fundamental modification which, although it still proved unsatisfactory from the point of view of reliable results and allowed no adequate time-weighting of the elements of complex activities, formed these elements into workable units and paved the way for the final technique.

This intermediate technique is described as follows:

Seven clearly defined types of activity were selected as most significant: (1) contact with material; (2) walking or other physical activity, not accompanied by contact with material; (3) absence of overt activity (chiefly looking, whether aimless or directed at persons or things); (4) talking; (5) physical contacts with persons; (6) laughing; and (7) crying.

A new record blank was devised, on which [the] duration to the nearest second was indicated in the appropriate column for the first three forms of activity. Concomitant activities were recorded on the same line, and each new activity was recorded on a separate line. It was soon found necessary to give up the attempt to time to the nearest second the last four types of activity, and their occurrence was merely indicated within the activities of longer duration. . . .

The focus of observation became the observed child, his overt and non-overt activity. Social contacts were defined objectively in terms of functional relationship only—that is, physical and verbal contacts. These activities, and instances of laughing and crying, were not timed, but were recorded once for each situation in which they occurred. In the case of talking and physical contacts, if the child made contacts with several different individuals during the same activity, the initials of each person were recorded in chronological sequence.[1]

The technique finally used in nursery school observations was readily derived from this stage by throwing all records of all

[1] Arrington, *op. cit.*, p. 20.

categories into five-second interval units and by dividing the total record form into two parts for simultaneous recording by two observers.

Two years of experimentation had been necessary to overcome the persistent optimism of the observers with regard to the amount of activity that one person could record at the same time. At the end of the experimentation period, however, it was clear that one person, however intelligent or carefully trained, could not record reliably a child's total activity—material, social, physical and emotional—at a given time, even when these were classified arbitrarily into the simplest objective units.[1]

III

Since the so-called Arrington technique represents in a way a culmination of our efforts to devise a satisfactory means of studying the basic social-material-self activity patterns of nursery school children, we have presented it with, perhaps, too little regard for the other observational studies in our social behavior program. Many of these were essential to this particular developmental stage and in themselves produced data with an important validating function for the data derived by this more abstract technique.

As we have described the derivation of the Arrington technique from the Barker technique, our change in procedure from *a posteriori* classification of activities to *a priori* definition has been emphasized. Our description seems to imply that this change was brought about somewhat as follows. The timed diary records failed to produce data which would fall reliably into our *a posteriori* classes, or which would differentiate children adequately on the basis of their behavior patterns. Therefore *a priori* observation classes were derived (a) partly through an analysis of the diary material collected by the Barker technique, (b) partly through logical analysis and (c) partly through successive experiments in recording. This does not, however, represent the whole of the process of definition as we actually carried it through. The selection of categories was also determined in large part by the experience gained from a number of other observational studies of discrete aspects of social behavior. It will perhaps illuminate the present stage of our investigations

[1] Arrington, *op. cit.*, p. 22.

further if we discuss briefly some of these other techniques and findings.

In the transition from the Barker to the Arrington technique, we had passed from a loosely defined concept of social contacts, in which each observer decided at the moment of observation whether a child was involved in a functional social relation with other children, or not, to a rigidly defined classification, in which the purely "spatial" element and the subtle gesture element in social interaction were eliminated and emphasis was placed on physical and verbal contact. Several studies, notably Hubbard's, Loomis,' Beaver's, Fisher's and Nelson's, contributed to the narrowing of this definition and to the validation of these data as representative of the more complex aspects of social interaction.

Hubbard's study[1] was contemporaneous with Barker's, but methodologically quite different. Her aim was to study the process of group formation by a random selection of all social situations in the nursery school. She attempted also to define the function of each child as regards the initiation and prolongation of the group activity. The technique was not successful (i. e., not reliable) for its main purposes, but it threw into focus some of the problems of definition.

A social situation was defined as a group of two or more children playing together, either functionally or spatially. This includes unitary activities such as a train game where six or eight children bring blocks and ride on the train. It also includes situations where two children play in the sandbox at the same time but where no observable coöperating occurs. It includes talking together about an airplane as the children look over the roof ledge, and it includes conflict over toys. But it does not include the social stimulation occurring when a child sees another with a toy and goes and gets another toy like it to play with. "Imitation" of that sort was not recorded as a social situation. Nor does the definition include children together spatially but playing with different toys, unless there is some objective evidence of interaction between the two. The exclusions are arbitrary, but were demanded by the exigencies of recording. . . .

While recording no attempt was made to classify the function [of each

[1] Thomas, Dorothy S., *Some New Techniques for Studying Social Behavior*, Child Development Monographs, No. 1. Bureau of Publications, Teacher College, Columbia University, 1929. Chap. IV. Hubbard, Ruth M., *A Method of Studying Spontaneous Group Formation.*

child] in such a way as leading, coöperating, objecting, etc. Instead, the actual rôle of the child in the group was stated in objective terms. For example, Dora was swinging, Eve was offering a toy to her; in another situation, Henry, playing with pans of water, said, "Get outa my house" and sent away other children, Adele played alone with her pan at the same bench, four other children were watching, and Harriet was trying to take Henry's pan away from him. Leadership, spectatorship, or routine activity in a group was to be classified later.[1]

It will be noticed, that the determination of what were differentia of functional social contacts was approached but not adequately solved. The arbitrary exclusion of certain spatial contacts diminished the subjectivity of decision but still depended largely upon *a posteriori* classification.

Beaver[2] first made an intensive diary record study of three small boys whose activities were so closely knit as to suggest a real "gang."

Though much of their social activity took place within this small group, many contacts were initiated with other members of the larger group with whom they were associated. All [their] contacts observed, whether "intra-gang" or "extra-gang," were recorded. We were interested in the factors involved in the formation of this group and the many subgroups. . . . Some . . . were in the nature of conversation. Others involved playthings or material of some sort; for instance, the child might appropriate the toy used by another. Still other contacts were merely physical, as a simple caress, a push, or a slap. Responses varied from an acceptance to a rebuff or refusal. In addition to differences in type of contact and response to contact, we found a wide diversity in number of contacts made, and the number of persons with whom they were made. One child might initiate many more contacts than another, but these contacts might be divided among several children. On the other hand, another child might initiate fewer contacts and all these contacts might be made with a single child. Thus we have two variables, one being the number of contacts initiated and the other the number of people with whom the contacts are initiated.[3]

On the basis of these diary records, which showed the part each member of the gang took in initiating contacts or in excluding outsiders, categories were derived for observational recording of the initiations and responses of a larger group.

[1] Hubbard, *op. cit.*, pp. 76–77.

[2] Beaver, Alma P., *The Initiation of Social Contacts by Preschool Children*, Child Development Monographs, No. 7. Bureau of Publications, Teachers College, Columbia University, 1932.

[3] Beaver, *op. cit.*, p. 3.

We found that nearly all initiations of social contacts involved a material or verbal element or both. For example, Tommy may exclaim as he grasps the handle of the doll buggy pushed by Alice, "I want it." This contact involves both a material and a verbal element. On the other hand, Tommy without preamble may reach unceremoniously for the buggy and pull it. This type of contact involves the material element alone. If, however, Tommy, through previous experience in such matters, has learned the language pattern approved by adults, he may say to Alice without making any attempt to touch her buggy, "When you are through with it, may I have it?" Here we have the contact with the verbal element alone. Still another type of contact occurs in which neither verbal nor material element enters; this category includes all caresses, blows, and touching of person or articles of clothing. The emphasis in our analysis is upon word or material rather than upon physical proximity or contact. Physical contact may or may not be present during any one of the different types of social contact mentioned. It is almost always present in the non-material non-verbal type. But even in this category a wordless gesture or beckoning may constitute a definite social initiation though the children be some distance apart. . . .

Certain accidental physical contacts which apparently have no social import occur in any group. We felt that these should not be included in our study. This made necessary the determination of criteria for differentiating accidental physical contacts from physical contacts socially significant. The procedure eventually fixed upon, stated negatively, is as follows: No physical contact which was apparently accidental was included if there was no follow-up by the child who made the contact, or no response, other than that which might be made by any inanimate object given the same treatment, from the child who was the object of the contact.

We find, then, four types of initiation of social contact which are mutually exclusive; namely, material verbal, material non-verbal, non-material verbal, and non-material non-verbal. [The following specific instances illustrate the meaning of our categories.]

[Initiations]

I. Material Verbal

a. Offering object or toy with verbal comment
b. Taking or grabbing object or toy with verbal comment
c. Bumping or pushing child while riding trike or kiddie kar with verbal comment
d. Pushing child in carriage or wagon with verbal comment
e. Striking child with object of some sort with verbal comment

. . .

II. Material Non-Verbal
- a. Offering toy or object without verbal comment
- b. Taking or grabbing object or toy without verbal comment
- c. Bumping or pushing child while riding trike or kiddie kar without verbal comment
- d. Pushing child in carriage or wagon without verbal comment
- e. Striking child with object of some sort without verbal comment

. . .

III. Non-Material Verbal
- a. Bumping into child with verbal comment
- b. Caressing child with verbal comment
- c. Touching person or clothing of child or adult with verbal comment
- d. Pulling or pushing child with verbal comment
- e. Slapping child with verbal comment
- f. Greeting child
- g. Any conversation which does not include an exchange of material

. . .

IV. Non-Material Non-Verbal
- a. Bumping into child without verbal comment
- b. Caressing child without verbal comment
- c. Touching person or clothing of child or adult without verbal comment
- d. Pulling or pushing child without verbal comment
- e. Slapping child without verbal comment
- f. Gesture of greeting
- g. Beckoning to another person

. . .

[Responses]

I. Acceptance
- a. Takes material object offered
- b. Gives up material object asked for
- c. Shouts or laughs in acquiescence when bumped or pushed
- d. Answers greeting
- e. Accepts invitation
- f. Responds verbally to conversation
- g. Follows out command or suggestion

II. Refusal
- a. Hangs on to material object, refusing to give it up
- b. Pushes away material object offered him
- c. Resists bumps, protesting and holding tricycle or other vehicle in position

   d. Resists physical contact, pulling away
   e. Protests verbally to any contact

III. No Response
   a. Person addressed apparently does not hear
   b. Contact utterly ignored
   c. Object of contact simply stares; no movement to step aside; no comment
   d. Child makes no more response than would any inanimate object given the same treatment[1]

Records of each child in two groups were taken in five-minute samples; every initiation by and toward the child under observation was recorded in terms of one of the four main headings, with an indication of the identity of the other person and the type of response in terms of one of three headings. Beaver's analysis is concerned with the differentiation of a group on the basis of their tendencies to begin, carry through and respond to contacts with other persons.

There is a wide range in number of contacts initiated by the children. The average number of contacts per five minutes varies from 1.62 to 22.61. The mean for the group is 9.56. . . .

Some most interesting contrasting tendencies are found in certain children when their responses to other children, and responses by other children to them, are considered. An inclination to withdraw from the group, or perhaps an indifference to the society of others and an absorption in self-activities, is indicated by the child who accepts few contacts, and resists or makes no response to many. The opposite tendency may be noted in the child who accepts a very high percentage of her contacts, and resists or ignores an insignificant number. The ineffectual type of personality may be found in the child whose initiations are constantly ignored.

There is great diversification among the group as to resistance to other children, and resistance aroused in other children. Some children resist many contacts, but seldom arouse resistance in others. The reverse may be true, or we may find a child not only arousing resistance in others, but resisting many contacts also. One small child, who had lived a pampered life previous to entering the nursery school, probably had never shared a toy. Quite naturally, therefore, the mere fact that another child was using a plaything that she desired did not deter her from trying to obtain it. The percentage of her resistance responses was high, for she refused to give up any plaything in her possession; but her percentage of resistance

[1] Beaver, *op. cit.*, pp. 4–9.

responses aroused was still higher, being more than double the mean percentage for the group.

One half the children in our nursery school group did not resist adults at all according to our records. The percentage of resistance responses to adults was not significant for the remainder of the group with the exception of four or five children. One of these, a small girl, aroused resistance in other children in 11 per cent of her initiations, a proportion much below the average for the group. She resisted the initiations of other children only 7 per cent of the time; this percentage was next to the lowest for the group. But we find her resisting adults 27 per cent of the time, the highest percentage for this category.[1]

Beaver's categories of the factors involved in the initiation of contacts were satisfactory in that they could be recorded reliably,[2] but unsatisfactory in that they did not differentiate clearly the activity-elements involved. She points out an apparent tendency for the verbal to increase at the expense of the material in an older group as compared with a younger group, but her technique suffers from the same difficulty as Barker's and Arrington's early classification-system. Her categories are complex, the elements involved cannot be made mutually exclusive and, therefore, appropriate quantitative weights cannot be assigned to them. She says, "The emphasis in our analysis is upon word or material rather than upon physical proximity or contact." But neither word nor material nor contact emerges in such a clear-cut way that their relative proportions can be indicated.

Loomis and Fisher studied, respectively, the physical contact and the word as elements in social interaction. Loomis[3] was primarily interested in developing a technique applicable to the genetic approach, i. e., one which would eventually show the degree to which physical contacts predominate as a means of carrying through social relations at an early age level and the extent to which they diminish in importance with the maturation of the individual.

In the early years of childhood, physical contact with another person seems to be an essential element in a large proportion of all social contacts

[1] Beaver, *op. cit.*, pp. 63–65.

[2] The same difficulty in determining reliability which is described for Loomis' technique (infra, p. 30) was found in this study.

[3] Loomis, Alice, *A Technique for Observing the Social Behavior of Nursery School Children*, Child Development Monographs, No. 5. Bureau of Publications, Teachers College, Columbia University, 1931.

made and received. This seems to be particularly true in that period before the child has grasped, to any appreciable extent, the use of language as a means of social initiation and response. He hits other children, excludes them from his activity by pushes and blows, takes their toys away from them. He touches them, explores their persons, their clothes, their toys. He pulls them into his activity, caresses them, offers them toys without regard to the other person's use for them. He brings forth physical responses from those with whom he is establishing contact. Adults take the objects he holds out, return, repulse, or ignore his caresses; associates of his own age pull back, hit him in return, flee from contact with him, caress him, or accept his advances passively. In this social interaction we find each child at times playing the rôle of initiator of these contacts and, again, responding to the contacts initiated by others. It is the purpose of this study to develop a technique which will make possible an accurate record of the physical contacts made and received by each of a group of children, to see whether individual differences in proportions of contacts made and received are clearly brought out, and to find whether individual children are characterized by the particular kind or quantity of their physical contacts.

From *a priori* reasoning the period in a child's life when a comparatively large segment of his social behavior is associated with physical contacts may be expected to come after he has learned to walk and before he is able to express himself fully through language and other means of communication developed as a result of maturation, experience, and training. The age following the learning to walk and preceding a rather extended use of language constitutes the nursery school span.[1]

Loomis studied several groups of children in the nursery school, taking a number of fifteen-minute records on each of the group. Every contact made and received by a given child during this period was indicated in terms of the identity of the other person involved, the rôle of the child observed, and the response to the contact. The specific categories were as follows:

Basis of Classification	*Category*
Rôle of person observed:	(1) subject or initiator of contact
	(2) object or recipient of contact
Classification of contact:	(1) accident
	(2) support
	(3) hit
	(4) push
	(5) pull

[1] Loomis, *op. cit.*, pp. 2–3.

Basis of Classification	*Category*
Classification of contact:	(6) caress
	(7) exploration
	(8) assistance
	(9) pointing
. . .	.
Kind of response:	(1) passivity
	(2) resistance
	(3) flight
	(4) coöperation[1]

The contacts were later combined into neutral contacts (classes 1 and 2), aggressive contacts (classes 3, 4 and 5), and coöperative contacts (classes 6, 7, 8 and 9). An index was also found for each child showing the ratio of the contacts made to contacts received (the so-called subject-object ratio).

The technique proved satisfactory (i. e., reliable) in differentiating children on the basis of the number of contacts made and the rôle of the child in the contact, but less satisfactory (unreliable) in differentiating among the descriptive categories. The extent of the differentiation of the children is indicated. For one group of 16 children, the range of the number of physical contacts in two hours was from 116 to 487, whereas the subject-object ratio varied from .38 to 1.46. There was a marked correlation between the total number of contacts and age, but the type of contact was apparently greatly modified with increasing use of the word as a means of initiating and prolonging social relations. The complexity of the interrelationship between the physical and verbal is indicated in Loomis' analysis of the statistical results in their context of available descriptive notations.

Roddy and Hendrick, who ranked high in both number of words and number of physical contacts, were the two older members of the "pre-school gang" described by A. P. Beaver. They were evidently much in contact with their associates. Hendrick who was in the main group of this study ranked high in both coöperative and aggressive advances and in the preponderance of initiated over received contacts—a manager in the making, it would seem. In fact, his ordering the other children about led to his receiving special educational attention from the teachers, thus probably introducing an influence associated with some of the irregularities in his record.

[1] Loomis, *op. cit.*, pp. 20–21.

Andrew and Oliver at the other extreme showed decidedly different characteristics from Roddy and Hendrick in that these two children had comparatively little association with other children through either language or physical contact. Age may have been one factor, for these boys were at or below the median age, and chronological age was found to be correlated with both number of words and number of contacts. It could not, however, be the only factor. The descriptive notes often referred to the active solitary play of these boys; and the fact that they had two of the three highest I. Q.'s in the group might be related to their ability to entertain themselves with relatively less association with other children. . . .

Of the three children showing a large negative relation between language and physical contact, descriptive notes of behavior were available only for Anna Marie. She frequently used language in such a way as to reduce the number of her contacts, as when riding a bicycle she asked a child to get out of her way. Her substitution of verbal for physical means of communication increased the proportion of physical contacts of which she was the recipient. The statistical record showed that she ranked next to the lowest in subject-object ratio and that half of the coöperation that she received was requested. An incident in the yard illustrates two characteristics found in her quantitative records, her use of language in requesting assistance and the low proportion of initiated physical contacts. Anna Marie was being drawn about in the wagon by Eugene and so was recorded as being the object of requested assistance. When Eugene wanted to "change about" and ride, a common practice which with most children, helped to keep the subject-object ratios near 1.0, Anna Marie got out of the wagon, told Eugene to get in, and as soon as he sat down said, "Now, you have had your turn. Get out and I'll ride." Eugene complied without a word.

These limited data on the number of words used by twenty nursery school children and the physical contacts made by them suggest that even at this age there may exist individual trends in the choices between these two means of communication with associates. On the other hand, a child may be consistent in his extreme, his average, or his limited use of both means of communication. This was the trend among half of the twenty children observed. This consistency in social contact may be associated with other characteristics—the two boys who rank high in both language and contacts tend to direct their associates' activities, and the two boys using relatively few words and making few contacts more often play alone and at times avoid their associates. A child may have a preference for one means of expression, and it is possible that this preference is a significant aspect of his personality.[1]

[1] Loomis, *op. cit.*, pp. 76–78.

In spite of Loomis' predominant interest in the changes in the behavior of the same children with maturation, her data on this point are very limited. It was impossible to obtain enough data to give valid statistical results. Only four children who had been part of the group studied in the nursery school in 1928 were available in kindergarten in 1929. Loomis' analysis of the behavior change in those four follows:

Larry as a three-year-old had ranked highest in number of contacts made and also in [calling forth] flights as responses to his advances; he had one of the highest subject-object ratios and the lowest rank in coöperative responses received. As a four-year-old, his conflicts were frequent and vigorous and he was so often isolated or carefully guided by the teacher that it was difficult to get an extended picture of his spontaneous activity.

Mildred in the nursery school was distinguished for her passivity and was the object of a comparatively large number of coöperative contacts from others. Her subject-object ratio of .3 was by far the lowest in the group. She had provoked only one resistant response during the observation, and the descriptive notes indicated that she often played the rôle of baby, receiving care and caresses from the larger girls. In the kindergarten Mildred played alone more than any other child. In one 15-minute observation she was three times the only child not playing in a group. She made many approaches to adults, once making nine in 15 minutes and at several other times leaving the playyard with a teacher. In fact, it was difficult to get an extended observation of her apart from the influence of teachers.

Winifred's only distinction statistically revealed in the nursery school was a high ranking for response by flight. The notes showed that her flights often were accompanied by crying, and that she frequently left her group when it became crowded, started playing in another place, was joined by others, and moved on as the group grew larger. . . . Her subject-object ratio was 1.1. In the kindergarten she exhibited the only complete change of behavior noted in the four children throughout the observation and this was only in the one trait of holding her own in every conflict. As before, she cried more than any other child in the room, initiated many plays, was always in a group, had comparatively little association with adults, and was not predominantly either the initiator or the recipient of contacts.

Nancy as a nursery school child had ranked nearly as high as Larry in number of contacts but unlike him she ranked highest among the eighteen children in coöperative responses received and very low in resistant responses. Her subject-object ratio, like Winifred's, was 1.1

and she made few contacts with adults. The descriptive notes told of her using language instead of physical contact in repulsing children who wished to join her play group. Although she received friendly advances from many children, she made them to only one child during the observation. A year later, in the kindergarten, she was found always playing in groups, though language was sometimes the only means of communication. At times she followed directions and at other times directed her associates, the conflicts were friendly, and adults were rarely in the picture. Her occasional objections to others joining her play group were expressed verbally and with courtesy.[1]

Fisher's study was based on reliable stenographic reports of the language of 72 children comprising from 6 to 12 hours consecutive recording of each child's language. Her aim was to separate the social, material and self connotations of the words used.

One of the main purposes of the study was to develop an objective technique for the analysis of children's language. The following categories for the purpose of analysis were developed and were found to satisfy the criteria of objectivity:

1. Grammatical form, independent of thought content, was used as an indication of the various stages through which children pass in gaining linguistic control.

2. The sentence, as one unit of vocal expression was used as a logical basis for consideration and analysis of the language. The data were recorded so that kinds of sentences used (simple, compound, complex, and compound-complex) appeared to indicate stage and complexity of sentence structure at different age levels.

3. Sentences were also analyzed . . . into four categories which indicated the degree of egocentricity and socialization of the child. The categories are:

I. Self as subject of sentence
II. Other person as subject of sentence
III. Thing as subject of sentence
IV. Non-verbal or incomprehensible remarks[2]

Our interest lay in the social differentiation, i. e., the relative proportions of speech that can be classified as being about the self, about other persons and about material objects.

Language may be primarily social in function, but any attempt to limit its purpose to communication distorts the developmental picture

[1] Loomis, *op. cit.*, p. 65–66.
[2] Fisher, Mary S., *Language Patterns of Pre-School Children.* Journal of Experimental Education, Vol. I, No. 2, pp. 70–71.

and blots out discriminating differences in individual patterns. From the point of view of child psychology perhaps the most important speech patterns are those which differentiate individual children with respect to their interests and desires.

The purpose of analyzing the present language records into the functional categories used has been to gain just such an index. The relative amount of time that children spend talking about themselves, compared to the amount of time they spend talking about other people or things, was roughly obtained by dividing sentences into three main categories depending upon the subject of the sentence. Such a division is obviously arbitrary, and doubtless often violates the intention or interest of the child. The justification for using the subject of the sentence as the criterion for the functional analysis lies in its objectivity and the desire to see what patterns, if any, would be indicated. . . .

In order to get some measure of true egocentricity and socialization of the child a coefficient of egocentricity was obtained for each subject. The proportion of egocentric remarks to total speech was obtained by dividing total remarks in Category I (Self as Subject) by the sum of total remarks in Categories II and III (Other Person as Subject; Thing as Subject). The mean coefficient for the entire experimental group was .53. This coefficient remained remarkably constant. That there was no relationship with increase in age in the pre-school years is indicated by a Pearson r of .0045. In other words, the pre-school children studied made a consistently high proportion of their total remarks about themselves at all age levels. They talked constantly about what they were doing from moment to moment.

The amount of speech was also considered a social index, and showed very definite developmental patterns. For the children studied the average number of remarks was ninety-two per hour. Marked individual differences were indicated by a range of 172 remarks per hour. There was no reliable sex difference. Amount of speech was positively correlated with age with a definite and steady increase up to the forty-eighth month. Then the amount remained fairly constant ($.56 \pm .05$). The correlation between IQ and amount of speech was not significant ($.20 \pm .07$).

The percentages of questions and commands in total speech also gave important developmental patterns as well as social and functional indices. For both questions and commands there was a steady increase in amount as age advanced, with a definite peak at about thirty-eight months. After that there was a slight falling off and a general leveling.

The proportion of negative sentences in total sentences increased as the children grew older. The correlation between chronological age and

per cent of negative sentences was .71 ± .04. Negation, as expressed in language, was greatest during the fourth and fifth years. No reliable relationship was found between intelligence and the amount of verbal negation. As soon as the word *no* was learned, it was found to have been used steadily and fluently throughout all the pre-school years. There was no significant increase in amount with advance in age, and no reliable sex difference. Individual differences, however, were marked.

The use of the first personal pronoun in the plural as an index of social development showed significant age differences (.72 ± .04). There was no significant relationship, however, between IQ and the use of *we*, *our* and *us* (.18 ± .07). The girls showed more rapid social development in this respect than did the boys insofar as use of *we*, *our* and *us* suggests increasing awareness of membership in a large group.

The amount of calling or mentioning children or adults by name was also analyzed. The amounts did not differ at different ages nor between the sexes. From the point of view of children's interests, however, it was significant that these nursery school children as a group talked to or about children more than twice as much as they talked to or about adults.[1]

Fisher's data also throw light on the stability of such indices of the verbal behavior of the same individuals from one year to the next.

As a rough estimate of the consistency or change of individual language patterns, a comparison has been made of data obtained on two samplings of the language of ten children taken at successive stages of development. At the time of the first sampling, these ten children were in the three-to-four-year-old group in the nursery school. . . . The following year second samples of their language were taken, by the same method and the same recorder, when they were in [a] kindergarten group. . . . The nature of the groups, the type of activities and daily routine, were enough alike in the two instances to warrant comparison of the samples.

The children were ranked according to per cents obtained on certain items of the functional analysis for both the 1929 and 1930 records, and then each child's earlier rank in a given item was correlated with the corresponding rank for the next year. . . .

Although no conclusions can sensibly be drawn from such a small number of cases, rather interesting tendencies toward consistency are noted. In the case of per cents of remarks in Category I (Self as Subject) and the coefficient of egocentricity, there is evidence of very close correspondence, since general maturation cannot have entered into these

[1] Fisher, *op. cit.*, pp. 72–73.

correlations as a causative factor. (It will be remembered that per cent of remarks in Category I and the coefficient of egocentricity are not positively correlated with age —r's = .0002 and .0045 respectively.)

The consistency in respect to per cent of remarks in Category II (Other Person as Subject) at two successive stages is strongly suggested. However, general maturation is doubtless a causative factor here, since there was a high positive correlation between per cent of remarks in Category II and chronological age. ($r = .69 \pm .04$). The same factor enters into any interpretation of the possible tendency towards consistency in the amount of speech and the per cent of sentences in total remarks.

. . .

Consistency in Language Patterns of Ten Children, shown by rank correlations between per cents obtained in 1929 and corresponding per cents for 1930

	r
Per cent of remarks in Category I	.81
Per cent of remarks in Category II	.73
Per cent of remarks in Category III	.26
Coefficient of egocentricity	.79
Per cent of questions in total remarks	.02
Per cent of commands in total remarks	.26
Per cent of sentences in total remarks	.64
Average remarks per hour	.48

The lack of apparent consistency in the proportions of questions and commands in total remarks at two stages of development is interesting in view of the fact that most of the second samples were taken after the ages when children tend to ask most questions and give most commands.

The only conclusion, therefore, that seems justified on the basis of these rough measures of individual consistency at successive stages is that a high degree of correspondence characterizes only those language patterns which imply egocentricity.[1]

Nelson's studies[2] wherein she attempted to derive some personality indices through records of the overt behavior of children during intelligence tests also had important implications for observational recording. She attempted to record for each child under examination every instance of:

1. Gross activity—defined as actually rising from the testing table to walk or run to another part of the room

[1] Fisher, Mary S., *Language Patterns of Pre-School Children.* Child Development Monograph, in press. Bureau of Publications, Teachers College, Columbia University.

[2] Nelson, Janet F., *Personality and Intelligence,* Child Development Monographs, No. 4. Bureau of Publications, Teachers College, Columbia University, 1931.

2. Use of test materials for purposes other than that indicated by the examiner—such as building a train or tower with the colored blocks or form-boards
3. Initiative in the test situation, subdivided into two classifications:
   a. Spontaneously repeating test just completed
   b. Reaching after or asking for test material other than that presented at a given moment
4. . . . spontaneous conversation—a simple remark, such as "I can't," "Where does this go?", "You tell me," "This is fun, isn't it?"—was considered one unit. Isolated words were also considered one unit, and the occasional more involved and complicated sentence was arbitrarily divided into clauses [each clause counting as one unit]

Spontaneous conversation was also subdivided into the following subgroupings:

a. Expressions of satisfaction
b. Asking for help
c. Asking for praise and reassurance
d. "I can't"
e. Comments on the immediate situation, such as "This is hard," "This is easy," "This is funny," etc.
f. Collective monologue—as defined and discussed by Piaget ("Talking to one's-self in front of others")
g. Incidental or miscellaneous
   (f and g—subsequently recognized as unreliable—led to a further classification: (1) relevant or (2) irrelevant to the immediate situation)[1]

Her classification was made according to categories defined *a priori*, and reliability (as measured by simultaneous recording) was fairly satisfactory for most of her classes but markedly unsatisfactory for the classes "relevant" and "irrelevant" conversation. She made a methodological contribution by comparing the observational classification made at the time the behavior occurred with a classification made afterwards from a stenographic report of the verbal aspects of this behavior. The equivocal nature of some of her results led her to discard as valid material for statistical indices all descriptive subdivisions of spontaneous conversation.

These various studies influenced the definition and delimita-

[1] Nelson, *op. cit.*, pp. 7–8.

tion of categories in the Arrington technique—(1) by clarifying the concepts of social initiation and response (2) by indicating what elements in social initiations and responses could be recorded with a sufficient degree of reliability to warrant their statistical manipulation and (3) by providing an independent and elaborated body of data with which the results obtained by the more abstract technique could be correlated, thus indicating the degree of significance of the latter in terms of the former.

Beaver's, Loomis' and Fisher's results (as described) were especially illuminating with regard to these points. Beaver pointed out the patterning of initiations in terms of the verbal and the material. Loomis indicated the predominance of physical contacts in the social relations of young children and the differentiating importance of the relative number of contacts made and received by a given child. Fisher showed the importance of language as a social behavior index.

All these studies showed, too, that an observational technique which can satisfy our primary requirement and produce data suitable for statistical use, must deal with extremely simple acts within the aspect of behavior which it is desired to study. They showed the present undesirability of applying statistical analysis to the more subtle aspects of human behavior because of the impossibility of eliminating the observer's varying preconceptions from his observations. That these simple acts have significance as indicators of the more subtle aspects is suggested (but, of course, not proven) by certain intercorrelations of the data obtained by several of these techniques. Loomis, for example, found a correlation of $-.73 \pm .09$ between contacts initiated by each of a group of children and measures of muscular tension,[1] suggesting (with a very limited amount of data, however) that relative ease in initiating social contacts of this sort may be an indication of psycho-physiological differentiation.

Loomis found little relation between the number of physical contacts and the number of words used during a given period ($r = .33 \pm .15$). This finding suggests that the two forms of social contact may complement each other, but also suggests the possibility that the egocentric nature of speech at this level may dominate and obscure the social. It was partly for this reason

[1] Loomis, *op. cit.*, p. 74.

that the Arrington technique allows for a clear differentiation between talking to the self and talking to another person. A third possibility is that the group correlation is lowered by the high negative correlation for a few individuals.

IV

The indices derived from Arrington's categories correlated highly with the relevant data from these independent investigations.[1] The correlation between the ranks of the children for Beaver's first and second categories (material-verbal and material non-verbal initiations) with the ranks of the same children for the number of intervals in which Arrington's records showed material activity occurring simultaneously with verbal or physical contacts was .84 ± .06. Ranks on Beaver's second and fourth categories (material non-verbal and non-material non-verbal initiations) correlated with Arrington's "subject" physical contacts by .82 ± .07. Beaver's verbal initiations, however, corresponded to Arrington's verbal contacts to a lesser degree, .55 ± .14.

A special study of the elements involved in the laughter of preschool children by Brackett[2] gave data on the relative frequency of laughter for part of the group, correlating with Arrington's frequency of laughter-intervals by .95 ± .02. (Sampling errors, however, were thought to have unduly inflated this correlation coefficient.)

The correlation between ranks for average number of remarks, as found by Fisher, and verbal ranks found by Arrington was .84 ± .06. Ranks on Fisher's category relating to remarks where things are the subject of the remark and Arrington's data on relative time spent manipulating materials was .86 ± .06; between Fisher's averages for remarks with another person as the subject and Arrington's average for initiated social contacts (physical and verbal) the correlation was .32 ± .19; but between Fisher's averages for remarks with self as subject and Arrington's average time spent in non-social activity the correlation was − .34 ± .19. Surprisingly, the coefficient was positive, although small, .30 ± .19, between remarks with self as subject and social

[1] Arrington, *op. cit.*, Ch. VI.

[2] Unpublished thesis.

contacts initiated, indicating, so far as the data go, that talking about the self is associated with the initiation of social contacts to an extent that obscures its function as monologue in non-social situations.

Although all these correlations are based on an inadequate amount of data, they lend a certain confidence to the use of the Arrington technique, and suggest (but do not assure) that the abstraction of these very simple elements of behavior will give valid indices of more complicated aspects of the same type of behavior and will define social-material-self activity patterns of the sort that our original hypotheses demanded.

Some assurance that the Arrington technique is defining activity patterns which are really characteristic of the individuals observed is given when her data are divided into sub-samples and correlations computed between corresponding samples. Two samples were formed, each comprising twelve five-minute records for each of twenty children. These records were numbered consecutively for each child, in the order taken, and the averages for each category on the "odd" records were correlated with the averages for each category on the "even" records, the child being the unit of comparison.

Pearson r correlation coefficients for the six types of behavior . . . were .69 ± .08 for physical activity, .69 ± .08 for talking, .58 ± .10 for contact with material, .56 ± .10 for physical contacts initiated, .43 ± .12 for no overt activity, and .30 ± .14 for physical contacts received. The small number of the records upon which the means were based probably affected the size of the correlations by increasing the effect of errors of measurement and of errors of sampling that were due to inconsistency in the behavior of individual children.[1]

Next, two similar samples were formed in terms of the average of sixty odd-minute observations and the average of sixty even-minute observations.

Correlation coefficients were .94 ± .02 for no overt activity, .94 ± .02 for talking, .93 ± .02 for contact with material, .91 ± .03 for physical activity, .78 ± .06 for contacts received, and .69 ± .08 for initiated contacts. These uniformly high coefficients are probably attributable to two factors—the larger number of samples upon which the means were based, and the comparison of consecutive minutes of activity in two out of every three of the sixty pairs of one-minute records.[2]

[1] Arrington, *op. cit.*, p. 61.

[2] Arrington, *op. cit.*, p. 63.

A final comparison by minutes, made on a random basis, included no two consecutive minutes and utilized two-thirds of the data in the previous sample.

For contact with material, r was .88 ± .04; for physical activity, .82 ± .05; for no overt activity, .82 ± .05; for talking, .78 ± .06; for initiated contacts, .41 ± .12; and for contacts received, .36 ± .13. Although the samples on which these correlations were based still involved a certain amount of carry-over of activity, when the same activity lasted for three or more consecutive minutes, and the first and third, or third and fifth minutes of the record were paired, the proportion was much smaller than for the odd-even sampling.[1]

Since all the coefficients but four from these several samples were greater than + .50 and no significant differences were found between the means of the comparable samples, there is indication of a fair degree of consistency in behavior patterns as defined by this technique. Since the classification of physical contacts was in terms both of the subject and the object of the contact, the consistency of this aspect of behavior could be further compared by correlating a given child's contacts on his own records, as one sample, with a second sample of his contacts as they appeared on other children's records.

The Pearson r coefficient of correlation between the direct and indirect observation ratios for the twenty children was .76 ± .06. This indicates that an aggregate of two hours of observation, distributed in five-minute samples over a period of months, gives a fairly adequate sample of a child's behavior with regard to the frequency with which physical contacts are initiated and received. This finding is substantiated by a Spearman rank-difference correlation of .79 ± .06 for subject contacts, and a correlation of .71 ± .08 for object contacts, obtained by pairing the ranks of the twenty children for average number of subject contacts per five-minute period on direct, and on indirect, observation, and their ranks for average number of object contacts on direct, and on indirect, observation.[2]

This correlation was more reliable than those based on the other samples, because of the longer observation period upon which the data for the indirect sample were based.

We have discussed in considerable detail the successive stages in the evolution of a recording technique designed to produce

[1] Arrington, *op. cit.*, p. 64.
[2] Arrington, *op. cit.*, pp. 68–69.

reliable data concerning activity patterns indicative of social-material-self interests in preschool children. That this technique is capable of defining activity patterns of individuals at this early age level with a high degree of certainty has been indicated by the results of the application of the best known statistical tests. The modifications necessary in techniques of this sort and the extent to which they are successful in defining characteristic behavior patterns of individuals at the kindergarten, adolescent and adult levels will be discussed in the three succeeding chapters.

CHAPTER II

SOCIAL-MATERIAL-SELF ACTIVITY PATTERNS IN A KINDERGARTEN GROUP

The free play period in the nursery school, described in the preceding chapter, represented an excellent situation in which to observe young children's spontaneous responses to an environment rich in material and social resources. Choice of play materials and of companions was unrestricted and no conformity to structuralized behavior patterns was expected. At the two-year level, adult interference with the child's self-chosen pursuits occurred only when physical needs demanded it. At the three-year level, game or story groups were occasionally initiated by adults, but the major part of the play activity remained essentially undirected. In the kindergarten, the next higher stage in the chronological scale, the daily program consists, in the main, of relatively fixed procedures to which all individuals are expected to react similarly. As in the nursery school, the lunch period, the rest period, the story and music periods, impose upon the child the necessity of behaving in certain definitely prescribed ways. The one situation which corresponds most closely to the nursery school play period, from the standpoint of minimal direction of activity and abundance of material and social resources, is the "work period" which occupies approximately the first hour of the morning. This period was selected, therefore, at the beginning of the study to be described here, as most suitable for the observation of the differential responses of children at the four-to-six-year level to the social, material and self components of their environment, and as most comparable to the factory situation in which the same aspects of behavior were being studied in adult workers.[1]

While this situation is freer from adult supervision of activity than any other in the daily program, we find in it the beginnings of structuralized behavior, to the extent that certain patterns are largely inhibited and others favored. Choice of occupation

[1] See Chapter IV, *infra*.

and selection of materials are still determined, with rare exceptions, by the child himself but it is understood that he is to work with material and that he is, if possible, to complete a piece of work during the period. He is not at liberty to favor his social interests to the complete exclusion of some functional use of materials. He must work at something, not wander aimlessly about as he might have been permitted to do in the nursery school. But since two teachers cannot watch continuously every individual in a group of 30–40 children, it is obvious that occasional violations of the work rule will occur. The child who wastes his work period, however, usually gets his public rebuke during the subsequent group discussion of work accomplished, and his tendency to avoid work is, to that extent at least, curbed.

The possibility of social interaction within this work situation is practically unlimited. Children are free to work where they please, to move about at will, to exchange materials, to change occupations and to converse freely concerning matters relevant or irrelevant to the work in hand. Certain equipment designed for play activity—dolls, toy furniture, toy trains, airplanes, etc.—is in the room, but, after the first weeks of the school year, it is understood that these materials are to be used only after some work has been completed. Freedom of self-expression is restricted only when a group becomes too boisterous. On rare occasions the work activity of the entire group is halted to impress the necessity for less noisy conversation or quieter use of materials.

The major purpose of the kindergarten project was similar to that of the nursery school studies. We hoped to discover, through consecutive short sample observations of the overt behavior of selected individuals, distributed over a period of months, what patterns of behavior toward materials, toward other persons and toward the self are generally characteristic of most individuals at a given age level, and to what extent specific individuals deviate from the general pattern. We were interested in studying consistency, as it is related to particular types of behavior and to the individuals manifesting the behavior. Studies of younger children had suggested that a given individual may show highly consistent patterning in some aspects of behavior and lack of patterning or inconsistency in others, that

another individual may show consistent patterning in all aspects of his behavior and that still another individual may be highly inconsistent and show no patterning in any of the aspects of his behavior which we are studying. The relative dominance in a total personality-complex of concern with inanimate material objects as contrasted with interest in social relationships, should be indicated by our data. Persistence as a personality trait can be estimated by the length of time spent in continuous use of materials and the number of breaks in the continuity of the work activity. One of the most interesting aspects of the personality picture upon which our data throw light is the relationship between an individual's outgoing and his incoming social contacts. On the active, "extrovert," side, we have his own efforts to assert his personality, to evoke responses from others, as measured in strictly behavioristic terms by the frequency with which he initiates verbal and physical contacts. A low frequency of outgoing contacts, on the other hand, emphasizes the passive, "introvert," side of the picture. A somewhat inadequate, but nevertheless suggestive, index of what Mark May has called the "social stimulus value" of a person for others (his ability to call forth responses from others with or without overt behavior on his part) can be derived indirectly from the records. Our aim, in short, in studying the relatively undirected activities of kindergarten children during the daily work period, was to obtain as comprehensive and varied, and, at the same time, as exact a body of data as possible concerning the social-material-self patternings of behavior in children at this age level.

The assumption underlying the nursery school technique was that, in an environment rich in social and material stimuli, the selective reactions of the individual as between materials and persons (or abstinence from such reactions) are significantly differentiating. This assumption, however, breaks down where the activities of the group are, to any marked extent, predetermined, i. e., where behavior patterns are already structuralized by the situation. As we have indicated, this was true in the kindergarten, where work was the specific focus of activity. The simple categories, *Material*, *Physical Activity* and *No Overt Activity*, lose their differentiating possibilities under these conditions, unless they are thrown into two larger classes, the par-

tially predetermined or structuralized *Job Activity* and the completely unstructuralized *Non-Job Activity*.

The job was broadly defined as an activity undertaken by a given individual, or individuals, in which functional contact with material was involved, and which resulted in a change in the form or the location of the material. While individual projects were the rule, and the achievement of a finished product by each child during the work period was encouraged, certain group projects were permitted. These job activities resulting in a group product could not be excluded from our study, since some children spent a large proportion of their time in directing, or coöperating in, group activities, rarely executing individual projects of their own. Activities defined as jobs in the 1931–1932 kindergarten observations may be classified broadly in two categories, the productive and the non-productive, productive being understood to mean 'resulting in an observable material product.' The first category includes the majority of the job activities,—drawing, painting, pasting, cutting, modeling, sawing, hammering, sewing, stuffing cloth dolls or other objects, stamping on paper, and building with blocks or dominoes. Each of these activities results in an observable product, a drawing, a painting, or a replica of some real object made out of paper, wood, clay, cloth, blocks or dominoes. The non-productive group includes dramatic play with dolls, doll dishes, doll carriages, toy telephones, flatirons, clothes lines, airplanes, trains, etc., and reading, or looking at, books,—activities which involve manipulation or transportation of material but have no observable end product. These activities, with the exception of the last-named, were arbitrarily excluded on the basis of their infrequent occurrence and their incomparability with other types of job activity. A few jobs do not fall strictly in either of the two categories which have been distinguished. They are productive in the sense that a product is achieved, but the product is predetermined by the nature of the job. This group includes picture puzzles and putting blocks or pegs in form-boards.

All jobs noted during the entire observation period were further classified according to qualitative similarity into the following relatively homogeneous categories,—(1) crayons,—drawing, coloring, with crayons (2) blocks,—building with blocks, playing

with dominoes and circus animals, making words with wooden letters (3) sewing,—stuffing cloth dolls, animals, etc. with cotton wool (4) cutting, pasting, etc.,—making valentines; cutting awning for house; making paper basket, hat, kite, pattern, flowers, etc.; pasting and cutting; stamping letters on paper (5) modeling with sand (6) painting (7) modeling with clay (8) drawing on blackboard (9) sawing, hammering, etc.,—making sail, firehouse, airplane, bed, boat, cart, box, with wood, cardboard, hammer, nails, etc. (10) picture books (11) picture puzzles.

It is clear from this general description of the situation in which records were taken that the observational method which had been developed for use in the nursery school could not be applied in the kindergarten without considerable modification. The first important change was the subdivision of the *Material* category of the earlier technique into two categories, *Job Material* and *Non-Job Material*. All materials used in the performance of a particular job were considered job materials. If, for example, the finished product were to be an Indian head-dress made of paper, the paper, crayons and scissors were job materials. If the end product were to be a doll's dress, the cloth, needle, thread, etc. were the materials classified as related to the job. Non-job materials, on the other hand, included all furniture, clothing (that of the individual observed or of others) and materials used by other persons, whether similar to, or different from, those used by the observed child. In group activities in which two or more children manipulated the same materials, as, for example, when two children looked at the same picture book, the material was classified under *Job* for both. The *Job Material* and the *Non-Job Material* categories included only activities involving active manipulation or transportation of material, that is, functional use of material.

Non-functional use of material was recorded under one of the two headings, *Job-Self* or *Non-Job Self*, according to whether it was related or unrelated to the work activity. Non-functional uses included (1) looking at material, (2) holding material for others to see, and (3) feeling, touching, or pointing to, material, without manipulating it. Using a pencil, crayon, paint-brush, chalk, etc., to draw pictures or using needles and thread to sew a doll's dress were functional. Twirling a pencil, waving paper in

the air aimlessly, fingering cotton wool while waiting for the teacher to return a doll that had been stuffed too full, holding or fondling a doll, even when the work had been defined as sewing a doll's dress, were considered non-functional. If the dress were tried on the doll, however, the use of the doll would become functional. Touching another child's dress was considered a non-functional use of non-job material, but adjusting or otherwise manipulating the dress was a functional use. Pushing a chair up to the table was functional, but supporting the hand on the back of the chair was non-functional.

It will be noted that the *Job Self* and the *Non-Job Self* categories in the revised technique have replaced the *Physical Activity* and the *No Overt Activity* categories in the technique developed for the nursery school level. The first of the last-mentioned classifications loses its discriminative value in the indoor work situation of the kindergarten, since the only physical activity possible is walking about the room. In the outdoor play situation of the nursery school the environment had been rich in equipment designed primarily for physical activity,—jungle gyms, walking boards, slides, etc.—as well as in abundant space for running and jumping. The *No Overt Activity* classification, which was defined as "standing, sitting, or lying still, making no observable reaction to the environment other than looking around, or looking at persons or things" has not lost its discriminative significance, but has been discarded in favor of the more comprehensive *Self* category, in the interest of a more logical classification of the total behavior of the individual.

Two additional categories, *Job Person* and *Non-Job Person*, had to be introduced to allow for the infrequent occasions when a child was in physical contact with persons, and was not using material. In instances of simultaneous use of material and contact with persons, the *Material* category was given precedence over the *Person*, since the personal contact was recorded on the simultaneous record of physical contacts.

In the continuous record of behavior made by one of the two observers, therefore, every activity was assigned to one of the six mutually exclusive categories,—*Job Material, Non-Job Material, Job Self, Non-Job Self, Job Person* or *Non-Job Person.*

Under *Job Material* was recorded all activity involving func-

tional manipulation[1] or transportation of job materials. Observation of material was used as a criterion of functional manipulation for all jobs except sewing, stuffing objects, and mixing paints or paste,—activities which could be carried on without continuous observation of the material. Functional transportation of job material was defined as carrying job-related material to a different location where it was subsequently used or left. If an object was merely carried in the hand while the child went to get another material, and was brought back to the original location without having been used, the activity was recorded under *Job Self* rather than under *Job Material* (e. g., going to closet to get paper to make pattern for fireman's hat, carrying toy hat later used as model, without looking at, or manipulating, it). Non-job material, when used for job purposes, became job material (e. g., using handkerchief to push needle through cloth). Removing paint or paste from the hands was recorded under *Job Material.*

Under *Job Self* was recorded all non-functional contact with job materials which was accompanied by overt behavior indicating concern with the job. This category included such activities as the following: (1) non-functional manipulation of job material while talking about the job, (2) holding job material and observing it, (3) holding job material for another person to observe, (4) holding job material while waiting for the teacher to look at it,[2] (5) holding job material when transportation of it was forcibly interrupted by an obstacle, (6) holding job material to prevent another person from taking it away, and (7) pointing to, or touching, job material without manipulation, using the fingers, or tools related to the job. Behavior was also classified in this category when there was no contact with job material, if the child was observing material, was listening to directions or conversation concerning his work, was talking about his work, was walking to get tools needed for work, or was waiting to ask the teacher a question about the job.

[1] Manipulation included all movement of material effected by the use of any part of the body (e. g. hitching a chair up to the table by a movement of the whole body without use of the hands was considered functional manipulation of non-job material).

[2] Time spent in waiting for the teacher's recognition or for the teacher to return job material was recorded under *Job Self* if the child indicated concern with the job by looking at the material or at the teacher, or by conversation; otherwise under *Non-Job Self.*

Under *Non-job Material* was recorded all behavior involving functional use of non-job material, such as opening and shutting closet doors, moving chairs, adjusting clothing, and manipulating, or transporting, material not related to the particular job in hand. Manipulation of job material which was extraneous to the end result of the job was recorded in the *Non-Job Material* column, but labeled job material (e. g., hiding drawings under table to prevent disliked child from seeing them).

The *Non-Job Self* category included all non-functional contact with job or non-job material, such activities as the following: (1) non-functional manipulation of material while looking at others, or while conversing about matters not related to the job, (2) holding job material and looking away from it, giving no overt evidence of attention to job activity, and (3) pointing to, or touching, non-job material (e.g., supporting self on chair or table, fingering clothing, rubbing hand on edge of table, twirling pencil, fingering crayons and looking away from drawing). Behavior was also recorded here when there was no contact with material and the child was walking about without any apparent destination, or was looking at other persons or at non-job material.

Under *Job Person* were recorded physical contacts with persons which were occasioned by the job, such as pushing a person away to prevent interference with work, or using physical contact as a means of attracting a person's attention to work. *Non-Job Person* included all other physical contacts, touching, embracing, hitting persons, etc., when there was no overtly observable relationship between the contact and the work activity.

The non-continuous record of language and physical contacts[1] made by the other observer was also revised to allow for the distinction between *Job* and *Non-Job*. Five categories were differentiated,—*Job Language*, *Non-Job Language*, *Indeterminate Language* (inaudible or otherwise unclassifiable language), *Job Contacts* and *Non-Job Contacts*. Speech which was distinctly heard by the observer was classified under *Job Language* when it referred to job materials, or to any process or idea pertaining to the performance of the job; under *Non-Job Language*, when

[1] The emotional aspects of behavior, laughing and crying, which had been included in th nursery school technique, were discarded as not sufficiently important to be included at the expense of a lowered reliability in the recording of other aspects. Crying, incidentally, was practically non-existent at the kindergarten level.

it pertained to any person, thing, or abstraction not related to the job. Non-verbal vocalization was classified under *Job* when it was obviously intended to represent the sound of a real object, a toy replica of which was being used as job material (the sound of a train, airplane, etc.); otherwise under *Non-Job*. Since the language of some children was indistinct, and since the observer frequently could not stand near enough to the child to hear his speech adequately without affecting the spontaneity of his behavior, it was necessary in many instances to infer the content of language from other overt behavior. Speech accompanied by pointing to, or demonstration of, job material, and speech addressed to the teacher, when the work activity had been interrupted for the purpose of consulting the teacher, were recorded under *Job*. Speech was recorded under *Non-Job* when the child was functionally occupied with non-job material, or when one job had been completed and another had not been undertaken. When the language was not heard and could not legitimately be inferred from other behavior, it was recorded in the *Indeterminate* category.

While no attempt was made in this study to record content of language, a few remarks noted incidentally by the observer when time permitted are given here by way of illustration of the *Job Non-Job* distinction applied to language.

Job Language

"I saw an airplane in the movies. It went that way." (Child flying a toy airplane.)
"I'm cutting the grass." (Child folding drawing representing grass.)
"That's the father flower." (Child explaining drawing.)
"Get out of my way."
"Don't talk to me any more."
"Don't bother me."
"I'm a good coloring man." (Child drawing with crayons.)
"How?" (Child asking another child how to do something.)
"I don't need it." (Child returns peg-board which he has tried to use in his block-building.)

Non-Job Language

"You're a hot cross bun."
"You're a piece of snow ball."
"What do we care?" (Other child telling what she is doing.)

"J-O-P J-O-P" (Child repeating names of letters heard at other table.)

"There are four dresses with green in them." (Child observing children at another table.)

"What?" (Child asking another person to repeat what has not been heard.)

"Do you want my crayons?"

"What color are you going to make it?"

"Do you know how to make a bunny rabbit?"

"Do you want to be a copy-cat?"

"Supposing it's pants?" (Child guessing what other child is drawing.)

"They don't go to school in summer. What do you think this world is going for, anyway?"

"My mother doesn't want any more children."

Physical contacts were differentiated according to whether they were related to the work activity or not, whether they involved material or not, and whether they were initiated or received by the individual observed. Each contact observed was recorded either in the *Job* or the *Non-Job* column, and was labeled with the appropriate symbol to indicate whether material was involved, and whether it was a subject contact (one initiated by the observed child), an object contact (one received by the observed child), or a subject-object[1] contact (one initiated simultaneously by both persons). Physical contacts with persons through material included only those involving manipulation or transportation of the same material by the observed child and others. Pointing to, or touching, another child's drawing paper when it was lying on the table was not considered a physical contact with the other child, but lifting it, or otherwise manipulating it while the other child was holding it, did count as a contact. Taking a crayon from a box held by another child did not count as functional, but taking the box out of the other child's hand was functional. Holding a valentine in place while another child drew around it was not recorded as a contact. Holding a board while another child hammered nails into it, unless the other child was at the same time manipulating the board, was nonfunctional. Casual contacts with a child made by the teacher, such as patting on the head, were arbitrarily omitted.

[1] This category was not distinguished at the nursery school level.

Contacts through material which were recorded under *Job* included exchange of materials related to the job (giving objects to, or taking objects from, persons), manipulation of material with persons, and transportation of material with persons. Job contacts not involving material included such activities as pushing a person away to prevent interference with work activity, touching a person to attract attention to the job, leading a person to the scene of job activity, and physical contacts with persons which were simultaneous with conversation about the job. Under *Non-Job*, contacts through material included touching persons with non-job materials (e. g., pushing a doll carriage against a person intentionally, pricking a person with a needle, hitting a person with scissors) and manipulation of material with which another person was in physical contact (e. g., pushing another child's peg-board on the table while the child was in contact with it, pushing another child's chair up to the table, turning the leaves of another child's picture book). Contacts not involving material, and classified under *Non-Job*, included all touching, embracing, hitting, pushing, etc., when there was no evidence of concern with work activity.

The following symbols were used in the recording of language and physical contacts.

Language

Δ—teacher
π—group (speech heard by the observer as definitely addressed to more than one person, or accompanied by observation of more than one person, or social vocalization, repetition of words or sounds spoken by others)
ss—self speech
s—singing
w—whistling
✓—non-verbal vocalization
X—speech heard by observer, identity of person addressed indeterminate

Initials were used to distinguish the different children in the group.

Physical Contacts[1]

MB—subject contact initiated by the observed child with Child MB

$\underline{\text{MB}}$—object contact received by observed child from Child MB

$\overline{\underline{\text{MB}}}$—subject-object contact initiated simultaneously by the observed child and by Child MB

$\left.\begin{array}{l} \text{MB}^{m} \\ \underline{\text{MB}}^{m} \\ \overline{\underline{\text{MB}}}^{m} \end{array}\right\}$—contacts involving functional use of material

The accompanying sample record of the block-building activity of Child Gr for the five-minute period beginning at 9:15 A. M. on May 11, 1932 (Chart VII, Form A), together with the simultaneous record of language and physical contacts (Chart VII, Form B), will illustrate the use of symbols.

Child Gr was occupied with job activity almost continuously throughout this five-minute period. Occasional brief interruptions occurred in the second, third, and fifth minutes. In intervals 26, 28 and 49 he used materials not related to the job. In Intervals 18, 19, 23, 34–36 and 60 he was distracted from job activity and made no contacts either with materials or with persons. His language was wholly related to the job. He talked to Child RD in four intervals, to himself in two intervals, to a group in three intervals, and to the teacher in two intervals. With one exception (Interval 57) his contacts were all occasioned by the job and all but two involved material.

It will be noted that, in addition to the changes in categories which have been described, some slight changes have been made in the form of the record blank.[2] A lengthwise blank which could be folded by minutes and held inconspicuously on a small wooden board containing the stop-watch had proved highly satisfactory in the factory situation, and was adopted in place of the more cumbersome blank previously used. To the time scale as used in the nursery school technique was added a consecutive numbering of the 60 five-second intervals composing the five-minute period and a numbering of the minutes.

Records of the type described were made daily in a public kindergarten having an enrollment of 35–40 children, the obser-

[1] A change in symbols used to differentiate initiator and recipient of contacts was necessary when the subject-object type of contact was added. Cf. pp. 2 and 4, *supra*.

[2] Cf. Chart I, p. 3, *supra*.

FORM A

CHILD Gr
DATE 5/11/32
TIME 9.15
ACTIVITY Big Blocks

TIME	NON-JOB			JOB		
	PERSON	SELF	MAT	MAT	SELF	PERSON
1-1-0						
2-5						
3-10						
4-15						
5-20						
6-25						
7-30						
8-35						
9-40						
10-45						
11-50						
12-55						
TIME	PERSON	SELF	MAT	MAT	SELF	PERSON
13-2-0						
14-5						
15-10						
16-15						
17-20						
18-25						
19-30						
20-35						
21-40						
22-45						
23-50						
24-55						
TIME	PERSON	SELF	MAT	MAT	SELF	PERSON
25-3-0						
26-5						
27-10						
28-15						
29-20						
30-25						
31-30						
32-35						
33-40						
34-45						
35-50						
36-55						
TIME	PERSON	SELF	MAT	MAT	SELF	PERSON
37-4-0						
38-5						
39-10						
40-15						
41-20						
42-25						
43-30						
44-35						
45-40						
46-45						
47-50						
48-55						
TIME	PERSON	SELF	MAT	MAT	SELF	PERSON
49-5-0						
50-5						
51-10						
52-15						
53-20						
54-25						
55-30						
56-35						
57-40						
58-45						
59-50						
60-55						

FORM B

CHILD Gr
DATE 5/11/32
TIME 9.15
ACTIVITY Big Blocks

TIME	VOCALIZATION			PHYSICAL CONTACT	
	JOB	NON-JOB	INDET	JOB	NON JOB
1-1-0				He^m	
2-5				He^m	
3-10					
4-15	RD				
5-20				RD^m	
6-25	RD			RD^m	
7-30	RD				
8-35					
9-40				RD^m	
10-45	RD				
11-50					
12-55				RD^m	
TIME	JOB	NON-JOB	INDET.	JOB	NON-JOB
13-2-0					
14-5					
15-10					
16-15	JJ				
17-20	JJ				
18-25	π				
19-30					
20-35					
21-40					
22-45					
23-50					
24-55					
TIME	JOB	NON-JOB	INDET	JOB	NON-JOB
25-3-0					
26-5					
27-10					
28-15					
29-20					
30-25					
31-30					
32-35					
33-40	Δ				
34-45	Δ				
35-50					
36-55					
TIME	JOB	NONJOB	INDET	JOB	NON-JOB
37-4-0					
38-5					
39-10					
40-15					
41-20				RD	
42-25					
43-30					
44-35					
45-40					
46-45					
47-50					
48-55					
TIME	JOB	NON-JOB	INDET	JOB	NON-JOB
49-5-0					
50-5	π				
51-10					
52-15				J^m	
53-20	π				
54-25					
55-30					
56-35					
57-40					RS
58-45					
59-50					
60-55					

CHART VII. SAMPLE FIVE-MINUTE RECORD OF THE "WORK PERIOD" ACTIVITIES OF CHILD GR

vations being distributed over a period of six months, from November, 1931 to June, 1932. The ten oldest boys and the ten oldest girls in the group at the beginning of the observation period were selected to form the observed group, since it was impossible to accumulate sufficient data on all children in the group in the time available. It even became necessary in the course of the study to drop four children from this group in order to obtain a minimum of 32 five-minute observations for each of the remaining 16 children. For one child who had several prolonged absences, it was possible to obtain only 24 records, and for another, who was transferred to another school during the observation period, only 30 records were obtained. The chronological ages of the children in this selected group ranged from five years and five months to six years when observations were begun. Observations were rotated in such a way that the records for each child were distributed as evenly as possible over the 8:30–9:30 work period, i. e., not made predominantly at one time during the period. No record was begun unless the child was actually manipulating or transporting job material, or had before him an unfinished product. If a child had completed one job and had not started another when his turn came to be observed, the next child on the list was substituted, and the first child's record was postponed until he had undertaken another job. If a job was completed during the five-minute observation, however, the record was continued until the end of the period. If the record was interrupted by the teacher's signal to the group to stop all work temporarily, indication was made on the record that an interruption had occurred and the record was begun again in the first interval of the minute in which the distraction had occurred. The child returned to the same job and was probably not seriously influenced by this type of distraction.

Analysis of the data contained in these 502[1] records of the highly spontaneous activities of 16 kindergarten children is still in an exploratory stage. The present discussion of results will, therefore, be limited to certain preliminary findings. The following points will be discussed briefly,—(1) the relative frequency of

[1] A group of 22 children of the same general age range is now being observed by the same method in another kindergarten. The number of records on this group already totals 600, and we hope to obtain 800–900 records before the study is completed.

different types of job for the group and for individuals, (2) the average group and individual frequency of the several kinds of behavior observed, (3) the relationship between frequency of occurrence of behavior and the kind of job, (4) the relative frequency of job and non-job activities in the behavior patterns of individuals, (5) the average length of periods of continuous use of job materials, as a measure of interest in materials and of persistence, and (6) consistency of the data in each aspect of behavior, as measured by correlations of sub-samples.

The relative frequency of different kinds of job per child and for the group as a whole is shown in Table I.[1] Drawing with crayons was by far the predominant activity for the group,—building with blocks, sewing, cutting, pasting, etc. coming next in order of frequency. Of the eleven kinds of job which we have differentiated, the average number in which each child was involved during the entire observation period was seven. Some interesting deviations from the group frequencies will be noted

TABLE I

Number of Five-Minute Records per Child Distributed According to Type of Job Activity

Child	Crayons	Blocks	Sewing	Cutting, Pasting, etc.	Sand Box	Painting	Clay	Blackboard	Sawing, Hammering, etc.	Picture Books	Picture Puzzles	Total Number of Records
1	9	1	..	1	6	..	4	10	..	1	..	32
2	5	11	..	2	5	3	2	..	1	..	3	32
3	16	3	..	1	1	1	3	..	2	..	3	30
4	16	..	6	..	..	2	..	..	..	..	..	24
5	19	..	5	3	..	1	2	1	..	1	..	32
6	9	5	4	2	5	4	1	..	1	..	1	32
7	11	7	..	1	2	1	2	..	1	4	3	32
8	8	7	3	3	3	5	..	1	2	..	..	32
9	19	..	..	3	..	1	..	5	..	4	..	32
10	14	11	..	2	2	..	..	1	..	2	..	32
11	22	..	4	3	1	1	..	..	..	1	..	32
12	15	1	7	4	2	1	1	1	..	..	..	32
13	28	..	..	2	..	2	..	..	..	..	..	32
14	17	..	9	1	1	2	1	..	..	1	..	32
15	6	8	5	2	1	2	..	1	7	..	..	32
16	12	4	..	3	1	..	5	1	3	1	2	32
Total	226	58	43	33	30	26	21	21	17	15	12	502

for individuals. Two children had an extremely narrow range of job activity. Child 4,[2] in 16 of her 24 records (67% of the time

[1] Each five-minute record was tabulated in the category corresponding to the job noted at the beginning of the observation, or, if a change of job occurred during the period, in the category corresponding to the job lasting the longer time.

[2] Numbers have been substituted in all tables and references for the initials used to distinguish children in the original records.

observed) had as her only job drawing with crayons. Sewing and painting were the only jobs undertaken in the other eight records. This child's prolonged absence may have been a factor in restricting the range of her activity with material. Child 13, for whom the maximum 32 records were obtained, also had a predominant number of records in which the job was drawing with crayons. In 28 of her 32 records (88% of the total time observed) crayons were the materials selected. In two of the remaining four, cutting and pasting was the activity, in the other two, painting. This child, judged by other indices to which reference will be made later, showed less interest in material than any other child in the group. At the opposite extreme is Child 6 who was involved in nine of the eleven kinds of jobs and distributed his time fairly evenly among them. He spent 28% of the time observed in drawing with crayons (nine records); 16% (five records), in block-building; 12½% (four records), in sewing; 6% (two records), in cutting, pasting, etc.; 16% (five records), in sand box activity; 12½% (four records), in painting; and 3% each (one record), in clay modeling, sawing and hammering, and in picture puzzles. Looking at picture books and drawing on the blackboard, the two types of job not recorded for this child, were activities usually selected by the lazy and inactive children in the group because there was little or no check on the end result of the activity. The three jobs selected by Child 13, it is significant to note, were activities which exacted little in the way of accomplishment. The distribution of records among a large number of job categories happens, in the case of Child 6, to be associated with high frequency of job activity. This is not necessarily the case, however, since Child 7 and Child 16, who also had records in nine of the eleven categories, rank respectively third and fourth lowest in the group in frequency of job material activity and second and fourth lowest respectively in frequency of total job activity.

While these children varied markedly in relative frequency of different types of job, we cannot interpret the figures as more than suggestive of possible preferences for certain materials on the part of different individuals, since choice of activity was not entirely free. Choice of job on a given day might be affected by availability of the material and of space at the table where the

material was being used, by the child's lack of equipment for a given job (e. g., painting and clay modeling, which necessitated apron or overalls), by the fact that the previous day's job had not been completed, or by direct suggestion from the teacher. The sampling of materials used by different children over a period of months was, however, probably not seriously distorted by factors such as these which occasionally restricted freedom of choice. Genuine preference for certain materials can perhaps be legitimately inferred, when a child has at least five more records of a given type than the group average for that kind of job. By this criterion Child 5, Child 9, Child 11 and Child 13 show a predominant preference for crayons; Child 2 and Child 10, for blocks; Child 14, for sewing; Child 1, for drawing on the blackboard; and Child 15, for sawing, hammering, and related activities.

The average distribution of time (average number of intervals per 60 intervals) among the six categories of behavior recorded on the material-self-person record is shown in Table II. As we

TABLE II

Average Number of Five-Second Intervals per 60 Intervals in which Job Material, Job Self, Job Person, Non-Job Material, Non-Job Self, and Non-Job Person Behavior Occurred in the Records of 16 Kindergarten Children

Child	Job Activity				Non-Job Activity			
	Material	Self	Person	Total	Material	Self	Person	Total
1	37.5	8.1	0.0	45.6	0.6	13.8	0.0	14.4
2	39.8	8.8	0.1	48.7	0.7	10.5	0.1	11.3
3	33.2	7.2	0.0	40.4	4.6	14.8	0.2	19.6
4	37.0	4.6	0.0	41.6	1.0	17.2	0.2	18.4
5	35.9	5.9	0.0	41.8	1.8	16.2	0.2	18.2
6	36.6	10.6	0.1	47.3	1.6	11.0	0.1	12.7
7	26.1	10.7	0.2	37.0	3.8	19.0	0.2	23.0
8	32.9	12.6	0.0	45.5	1.4	12.8	0.3	14.5
9	31.8	9.4	0	41.2	1.6	17.0	0.2	18.8
10	25.5	17.5	0.1	43.1	1.6	15.0	0.3	16.9
11	36.6	7.0	0.0	43.6	1.7	14.7	0.0	16.4
12	33.8	8.3	0.1	42.2	1.9	15.8	0.1	17.8
13	24.9	5.8	0	30.7	2.4	26.8	0.1	29.3
14	33.5	5.8	0.0	39.3	3.3	17.3	0.1	20.7
15	35.2	7.4	0.1	42.7	3.4	13.8	0.1	17.3
16	29.6	9.8	0.1	39.5	2.1	18.3	0.1	20.5
Group Average	33.1	8.7	0.1	41.9	2.1	15.9	0.1	18.1

would expect, from the definition of the situation in which records were taken, work activity or work-related activity occupied the predominant proportion of time in each five-minute period. The group average for *Total Job* was 41.9 intervals out

of 60, for *Job Material*, 33.1 intervals, and for *Job Self*, 8.7 intervals. The largest proportion of non-job activity falls in the *Non-Job Self* category, the group average being 15.9 intervals out of 60. Functional use of materials not related to the job was rare, occurring on the average in only 2.1 intervals per 60. Amount of behavior recorded under *Job Person* and *Non-Job Person* was negligible (the group average for each of these categories being 0.1 intervals). Individual children varied considerably from these group averages. The range for average number of intervals of *Total Job* per child was from 30.7 to 48.7; for *Job Material*, 24.9–39.8; for *Job Self*, 4.6–17.5; for *Total Non-Job*, 11.3–29.3; for *Non-Job Material*, 0.5–4.6; and for *Non-Job Self*, 10.5–26.8. Child 2 is the extreme example of the work-preoccupied child, having the highest average frequency of functional use of material and of total job activity, while at the other extreme is Child 13 whose interest in material, as judged by these same measures, is the lowest for any child in the group. The job activity of Child 2 was distributed among eight of the eleven categories (slightly more than the average distribution for the group) whereas Child 13, as previously noted, used the same materials in all but 4 of her 32 records.

TABLE III

Average Number of Five-Second Intervals per 60 Intervals in which Language and Physical Contacts Occurred in the Records of 16 Kindergarten Children

Child	Language						Physical Contacts							
	Job	Non-Job	Ind.	Social	Non-Social	Total	Job	Non-Job	Subj.	Obj.	Subj.-Obj.	Mat.	Non-Mat.	Total
1	8.4	4.8	1.7	14.1	0.8	14.9	0.7	0.2	0.3	0.5	0.1	0.4	0.5	0.9
2	7.0	3.0	1.9	10.7	1.2	11.9	3.2	0.7	1.2	1.3	1.4	3.0	0.9	3.9
3	7.1	8.7	0.8	15.6	1.0	16.6	1.1	1.7	1.9	0.4	0.5	1.2	1.6	2.8
4	3.2	8.8	1.3	11.5	1.8	13.3	0.3	0.8	0.6	0.4	0.1	0.3	0.8	1.1
5	3.2	5.1	0.6	8.0	0.9	8.9	0.3	1.0	0.5	0.7	0.1	0.7	0.6	1.3
6	10.7	3.2	1.0	14.1	0.8	14.9	1.6	0.4	0.7	0.6	0.7	1.0	1.0	2.0
7	7.1	10.3	1.2	17.9	0.7	18.6	1.6	1.7	1.7	1.1	0.5	1.4	1.9	3.3
8	5.8	3.5	0.5	9.3	0.5	9.8	1.2	1.4	0.4	1.4	0.8	0.8	1.8	2.6
9	5.9	4.0	0.7	10.2	0.4	10.6	0.9	0.7	0.3	1.0	0.3	0.5	1.1	1.6
10	6.9	5.9	1.3	11.3	2.8	14.1	1.4	0.9	0.7	1.2	0.4	1.2	1.1	2.3
11	3.9	5.1	1.2	8.6	1.6	10.2	0.3	0.3	0.3	0.1	0.2	0.3	0.3	0.6
12	9.0	12.3	1.6	19.9	3.0	22.9	1.1	1.2	1.0	0.8	0.5	1.0	1.3	2.3
13	2.7	7.3	0.9	9.5	1.4	10.9	0.2	0.9	0.4	0.5	0.2	0.4	0.7	1.1
14	5.9	8.1	0.6	11.8	2.8	14.6	0.7	0.9	0.7	0.4	0.5	0.9	0.7	1.6
15	10.4	7.6	0.9	16.2	2.7	18.9	1.6	1.7	1.4	0.7	1.2	2.1	1.2	3.3
16	8.2	9.3	0.6	16.6	1.5	18.1	1.0	0.7	0.2	0.9	0.6	1.0	0.7	1.7
Group Average	6.6	6.7	1.0	12.8	1.5	14.3	1.1	0.9	0.8	0.7	0.5	1.0	1.0	2.0

The average frequencies of language and physical contacts per

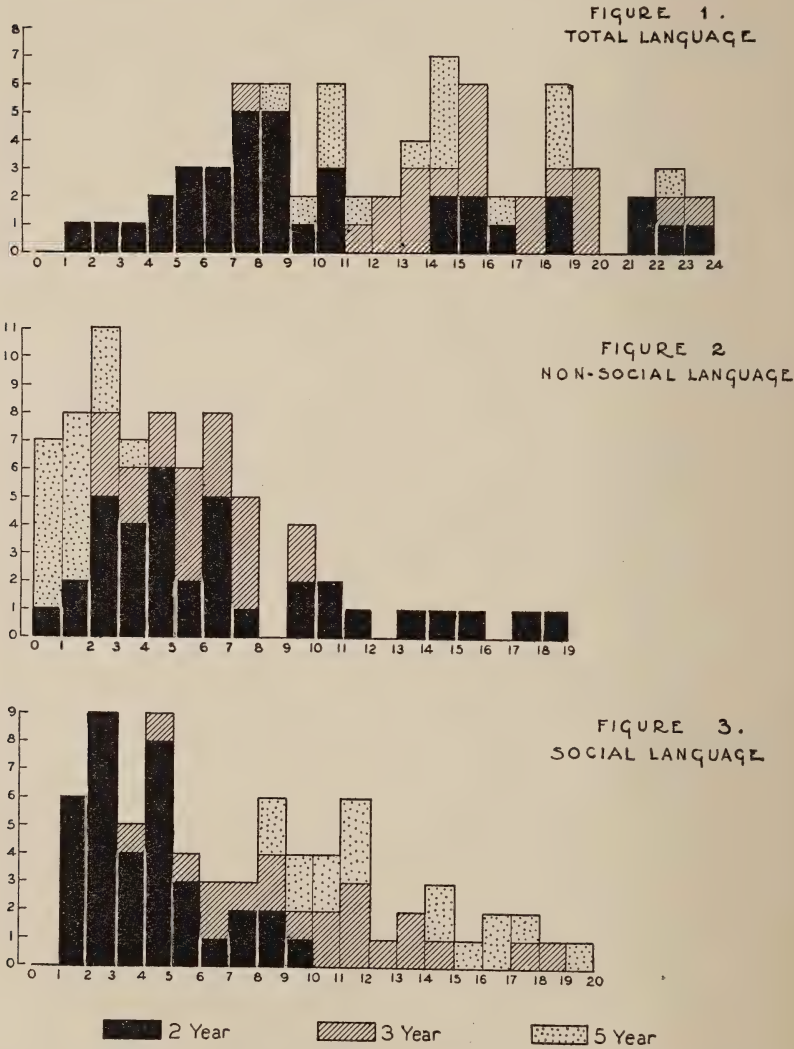
child and for the group are shown in Table III. The average group frequency of *Total Language* was 14.3 intervals per 60. Group averages for *Job* and *Non-Job Language* were practically identical (6.6 and 6.7 intervals out of 60, respectively), and 1 interval in 60 was classified as *Indeterminate*. *Non-Social Language* occurred, on the average, in 1.5 intervals out of 60, and *Social Language*, in 12.8 intervals.[1] The range of frequency for individuals within the group was, for *Total Language*, from 8.9 to 22.9 intervals; for *Job Language*, 2.7–10.7 intervals; for *Non-Job Language*, 3.0–12.3 intervals; for *Non-Social Language*, 0.4–3.0 intervals; and for *Social Language*, 8.0–19.9 intervals.

Since the method of recording language in the kindergarten study was exactly comparable to that used in previous studies of younger children, a graphic comparison of groups of children at the 2–3 year, 3–4 year, and 5–6 year levels is shown in the accompanying histograms. The average frequencies of *Total Language* (average number of intervals in 60 in which vocalization of any kind occurred) for the 36 two-year-old children tend to be concentrated, as we would expect, at the low frequency end of the scale. Some individuals in this group, however, have frequencies as high as those for children in the two older groups. (Fig. 1). The frequencies for the 20 three-year-old and for the 16 five-year-old children are distributed over approximately the same range. This overlapping may mean that language expression reaches a peak within the 3–4 year age range that is not exceeded at later ages, or that the more restricted conditions of the kindergarten work period inhibited to a certain extent the spontaneity of linguistic expression, or possibly that the composition of the groups at the different age levels was unrepresentative of "normal" language behavior at these ages. The decrease in non-social vocalization, and the increase in frequency of speech addressed to persons, with increase in chronological age, are clearly shown in Figures 2 and 3 respectively.

The average occurrence of the physical type of social contact for this group of kindergarten children was very low, such contacts occurring in only 2 out of 60 time intervals (Table III).

[1] Social speech included all language addressed to persons whether the identity of the person addressed was known or not. Non-social language included self speech, singing, whistling, and other forms of vocalization not definitely addressed to other persons. Group singing, repetition of rhymes with others, etc., were classified as social vocalization.

AVERAGE NUMBER OF TIME INTERVALS PER 60 IN WHICH TOTAL LANGUAGE, NON-SOCIAL LANGUAGE, AND SOCIAL LANGUAGE OCCURRED IN THE RECORDS OF 36 TWO-YEAR-OLD, 20 THREE-YEAR-OLD AND 16 FIVE-YEAR-OLD CHILDREN



The average frequency of job-related contacts was slightly greater than that of contacts irrelevant to the job (1.1 and 0.9 intervals per 60, respectively), the frequency of contacts initiated

(subject) was practically identical with that of contacts received (object), 0.8 and 0.7 intervals, respectively, and the averages for material and for non-material contacts were identical. The range of variation in individual averages was narrow,—for total contacts, 0.6–3.9 intervals; for job-related contacts, 0.2–3.2; for contacts not occasioned by the job, 0.2–1.7; for contacts initiated by the observed individual, 0.2–1.9; for contacts received by the observed individual, 0.1–1.4; for contacts involving material, 0.3–3.0; and for those in which no material was involved, 0.3–1.9. Child 2, whose preoccupation with work activity has been noted, had the highest average frequency of contact related to the job and of contacts involving material.

Comparisons with data for children at the two-to-four year level, such as were presented for language, are not given for physical contacts, for the reason that certain changes in definitions, made effective in the kindergarten study, undoubtedly affect the comparability of the two sets of data. Such changes were the addition of the subject-object type of contact, the elimination of certain kinds of contact, such as "being used as a means of support, or vice versa," and delimitation of the definition of contacts through material.

We have compared the relative frequencies of different kinds of behavior in the average five-minute sample, without regard to the kind of job in connection with which the behavior occurred. Since the nature of the job differed considerably according to the materials used, it is important to consider also the extent to which frequency of occurrence of specific behavior patterns is related to the job. To establish with any certainty relationships between occurrence of behavior and kind of job, we would need to rule out the influence of the individuals performing the job. It would be necessary, in other words, to have an even distribution of the records of the particular kind of job activity among all, or most, individuals in the group studied. This ideal situation obviously does not occur in our data. For only one of the 11 job groups do we have records for all of the 16 children, and, for five job categories, there are records for only half, or less than half, of the group. (Table I). Even though the distribution of jobs among the group is irregular, a comparison of average frequencies of behavior per category of job activity may be of

interest as suggesting possible relationships and as a background for conclusions concerning frequencies of behavior in individual patterns.

The average frequency of occurrence of the eight major categories of behavior for each of the eleven job groups is shown in Table IV. A high frequency of non-job self activity occurs in

TABLE IV

Average Number of Five-Second Intervals per 60 Intervals in which Different Kinds of Behavior Occurred in Records of Different Types of Job Activity

Behavior Category	Job Category											All Job Categories Combined
	Crayons	Blocks	Sewing	Cutting, Pasting, etc.	Sand Box	Painting	Clay	Blackboard	Sewing, Hammering, etc.	Picture Books	Picture Puzzles	
Job Material	29.9	28.3	37.1	36.7	45.1	39.7	47.7	33.9	31.8	18.8	39.8	33.1
Job Self	7.0	18.4	4.8	10.0	8.1	6.3	2.7	9.2	9.4	22.1	5.6	8.7
Job Person	0.0	0.2	0	0.1	0.2	0	0	0.0	0.2	0.0	0.1	0.1
Non-Job Material	2.0	2.2	3.0	2.4	0.8	2.1	1.2	1.0	3.4	1.2	5.2	2.1
Non-Job Self	21.0	10.7	14.8	10.8	5.8	11.8	8.2	15.8	15.1	17.4	9.0	15.9
Non-Job Person	0.1	0.2	0.3	0.0	0.0	0.1	0.2	0.1	0.1	0.5	0.3	0.1
Language	13.1	18.5	15.6	13.2	19.7	8.4	21.8	8.3	14.4	12.7	14.8	14.3
Physical Contacts	1.3	6.1	1.4	2.0	1.7	1.1	1.0	1.0	3.4	1.3	4.6	2.0

records in which the job was drawing with crayons (the average frequency per record being 21 intervals in 60, as compared with 15.9 intervals, the average per record for all jobs combined). This finding is explained by the fact that crayon drawing was accompanied by frequent observation and discussion of other children's drawings. In records of block-building activity, we find a relatively low frequency of job material activity (28.3 intervals as compared with 33.1 for all jobs combined), a high frequency of job self activity (18.4 intervals as compared with 8.7 for all jobs), a relatively low occurrence of non-job self activity (10.7 as compared with 15.9), a fairly high language frequency (18.5 as compared with 14.3), and the highest frequency of physical contacts found for any job category (6.1 as compared with 2.0). This job involved a large amount of observation of material, of conversation relevant to the job, and a

frequent exchange of job materials involving physical contact. Records in which the job was modeling with sand show a high frequency of functional use of job material (45.1 intervals as compared with 33.1), a low frequency of non-job self activity (5.8 intervals as compared with 15.9), and a high proportion of language (19.7 as compared with 14.3). This job was usually a group project, and involved almost continuous manipulation of job material accompanied by conversation dealing predominantly with job activity. For this job and for activity with blocks the average frequency of job language was approximately twice as great as the average for all jobs combined. Modeling with clay, on the other hand, was accompanied by a frequency of non-job language almost twice as great as the average for all jobs combined. Records of clay modeling, however, have the highest frequency of behavior classified as job material (47.7 intervals as compared with 33.1), with correspondingly lower frequencies in the *Job Self* and *Non-Job Self* categories, and the highest frequency of language (21.8 as compared with 14.3) noted for any job group. On the average, four out of every five minutes of this type of job were spent in manipulation of job materials, and conversation was more often irrelevant than relevant to the job. Painting and drawing on the blackboard were accompanied by low frequencies of language (8.4 and 4.3 intervals, respectively, as compared with 14.3 for all jobs combined). Looking at picture books involved the lowest frequency of job material activity and the highest frequency of job self activity noted for any kind of job. The frequencies of non-job material behavior and of physical contacts for the job activity groups 9 and 11 probably reflect the characteristic behavior of individuals rather than inherent characteristics of the jobs themselves. This summary of some relationships between behavior and kind of job which are indicated by the data may serve to emphasize the importance of relating findings for individuals to group averages for particular jobs.

As a further means of determining whether the frequency of a given form of behavior is associated with the job itself or with the individuals performing the job, average frequencies of behavior per job category might be computed for each individual separately. This was done for the physical contact category. The high

frequency of contacts which was found to be associated with block-building activity appears to be a genuine concomitant of this kind of job, since all of the ten children who used blocks have physical contact averages for this job that are as high as, or higher than, the average frequency of contacts for all jobs combined. For all but three children the average is distinctly higher than that for all records regardless of the kind of job. The job category having the next highest average frequency of physical contacts, picture puzzles, includes records for only five of the sixteen children but four of these five children have averages for this kind of job that are considerably higher than the average frequency of contacts per record regardless of kind of job. The high average for Category 9,—sawing, hammering, and related activities,—is overweighted by an extremely high frequency for Child 7 in one record.

As possible indices of important personality differences, ratios of job to non-job language, of job to non-job contacts, of material to non-material contacts, and of subject to object contacts were computed for each child, and are presented in Table V. The

TABLE V

Ratios of Job Language to Non-Job Language, Job Contacts to Non-Job Contacts, Material Contacts to Non-Material Contacts, and Subject Contacts to Object Contacts, for 16 Kindergarten Children

Child	Language	Physical Contacts		
	Job / Non-Job	Job / Non-Job	Material / Non-Material	Subject / Object
1	1.75	2.73	.75	.65
2	2.32	4.14	3.20	.98
3	.81	.63	.73	3.26
4	.37	.44	.30	1.36
5	.64	.32	1.16	.67
6	3.32	4.42	1.10	1.10
7	.69	.91	.78	1.39
8	1.62	.91	1.42	.49
9	1.48	1.38	.39	.33
10	1.16	1.55	1.00	.60
11	.76	.82	1.00	1.86
12	.73	1.00	.80	1.24
13	.38	.25	.67	.94
14	.74	.70	1.43	1.62
15	1.37	.89	1.74	1.42
16	.88	1.35	1.45	.44
Group Average	1.00	1.14	1.01	1.00

group averages show an even balance between the total number of occurrences of work-relevant and of work-irrelevant language, between the number of contacts involving material and the

number not concerned with material, and between the number of contacts initiated and the number received, but an appreciable excess of physical contacts relevant to the job over non-job contacts. For individual children, however, the size of the ratio and the relationship between different ratios vary considerably. Of the seven children whose ratios indicate an excess of language related to work activity, five have also an excess of work-related contacts, and of the nine children who have less job than non-job language, seven have also fewer job than non-job contacts. Child 8 and Child 15 have an excess of job-related language but not of job contacts. Child 16 has an excess of job contacts without a corresponding predominance of job language, and Child 12 has as many job as non-job contacts but has less job than non-job language. Since, by definition, job contacts might or might not involve material, a high job non-job contact ratio does not necessarily imply a high material non-material contact ratio. To interpret the significance of these indices by specific examples, Child 1 tends to talk more often about her work than about other matters, to make contacts of the work type more often than of the non-work type, to make contacts without, more often than with, material, and to be the recipient more often than the initiator of physical contacts. Child 2 differs in one respect, the tendency to be involved predominantly in physical contacts of the material type. Child 3 has the largest excess of initiated over received contacts and his language and contacts are more often unrelated than related to his work. Child 6 also makes more contacts than he receives, but he has the largest ratios of job to non-job language and of job to non-job contacts of any child in the group and his contacts are more often of the material than of the non-material type. Child 13, whose lack of interest in work activity has been noted, ranks low in all four indices. She tends to be the recipient rather than the initiator of contacts, and her speech and contacts are more often unrelated than related to her work.

As indices of interest in materials and of persistence as a personality characteristic, the average and the median length of periods of continuous job material activity were computed for each child. Since use of job materials was frequently in process when the five-minute record was begun and often continued

after the observation had been completed, it is obvious that some of the periods of continuous job material activity shown in the records are potentially incomplete.[1] While the exact length of these periods is indeterminate, we know that they are either as long as, or longer than, they appear in the records, i. e., the error lies in the direction of underestimation of the actual length of period. Exclusion of these indeterminate periods reduced the amount of data appreciably and lowered the means for all children to some extent. It is probable, therefore, that means based on the total number of periods, complete and incomplete, constitute the most representative measure we can obtain from our data, but they undoubtedly underestimate the actual length of period systematically for all individuals. Presumably no extreme measures have been excluded by this method, however, since there is no recorded instance of a period of behavior as long as 60 five-second intervals and only 19 out of a total 6230 periods (0.3%) lasted as long as 24 five-second intervals.

TABLE VI

Number of Periods of Continuous Job Material Activity per Child Distributed by Number of Intervals in Period, with Mean and Median Number of Intervals in Period

Child	Number of Periods of Continuous Job Material Activity Lasting							Total Number of Periods	Mean Number of Intervals in Period	Median Number of Intervals in Period
	0–5.9 Int.	6–11.9 Int.	12–17.9 Int.	18–23.9 Int.	24–29.9 Int.	30–35.9 Int.	36–41.9 Int.			
1	350	45	10	3				408	2.9	2.0
2	369	41	13	2	1	1	1	428	3.0	1.7
3	338	36	6	2	1			383	3.6	1.7
4	241	34	3	4	1			283	3.1	2.0
5	291	37	12	4	2			346	3.3	2.1
6	311	36	12	2	3			364	3.2	1.9
7	395	21	2	1		1		420	2.0	1.4
8	309	33	6	2	1	2		353	3.0	1.8
9	374	25	11	3				413	2.5	1.6
10	412	21	2					435	1.9	1.5
11	369	33	8	2	3			415	2.8	1.9
12	387	29	10	4				430	2.5	1.6
13	420	9	5					434	1.8	1.4
14	311	37	7	5				360	3.0	1.8
15	327	45	7	2	1			382	3.0	2.0
16	339	27	7	2		1		376	2.5	1.6
Group Average	346.4	31.8	7.6	2.4	0.8	0.3	0.1	389.4	2.7	1.8

Table VI shows the distribution of the total number of job material periods per child by length of period (number of con-

[1] The proportion of potentially incomplete periods varied, for 11 of the 16 children, from 9 to 11% of the total number of periods, and, for the remaining five children, from 6 to 8%.

secutive time intervals in which job materials were used functionally) together with the mean and median length of period. The usual length of periods in which work materials were used continuously was extremely short for the group as a whole, the mean number of intervals being 2.7, the median, 1.8. On the average, 346.4 of a total 389.4 periods (89%) lasted less than 6 five-second intervals, i. e., less than a half minute. The mean length of period varied from 1.8 to 3.6 intervals, the median length, from 1.4 to 2.1 intervals, for different children. Of the 434 periods spent by Child 13 in use of job materials, only 14 (3%) lasted a half minute or longer. At the other extreme is Child 5 who had 55 out of a total 346 periods (16%) lasting as long as, or longer than, half a minute. A more meaningful index of concern with the job may be the average or median length of periods of continuous job activity (*Job Material, Job Self* and *Job Person* combined). This point remains to be investigated.

In order to test the consistency of individuals in different aspects of behavior and the consistency of unrelated samples of the data concerning these aspects[1] (i. e., the reliability of the data), the five-minute records for each child were numbered consecutively in the order taken, and the mean number of intervals of behavior per five-minute sample for all odd-numbered five-minute records per child combined was correlated with the mean number of intervals for all even-numbered records combined. Pearson correlation coefficients for all categories except *Job Person* and *Non-Job Person*, in which the amount of behavior recorded was negligible, are shown in Table VII. The coefficients for total job activity and for total non-job activity (.72 ± .08) are necessarily identical, since these two categories complemented each other, the sum of the means per child in odd samples of job and non-job activity totaling 60 intervals, and the sum of the means per child in even samples of job and non-job activity also totaling 60 intervals. The categories of job-related activity show, in general, greater consistency than the non-job categories. The coefficient for *Job Material* was .75 ± .07, for *Non-Job Material*, .51 ± .12, for *Job Self*, .85 ± .05, for *Non-Job Self*, .69 ± .09, for *Job Contacts*, .88 ± .04, and for *Non-Job Contacts*,

[1] These two factors cannot be clearly distinguished at the present stage of development of our method.

TABLE VII

Pearson Coefficients of Correlation between Mean Number of Intervals of Behavior per Child for 16 Odd-Numbered Five-Minute Samples and Mean Number of Intervals of Behavior per Child for 16 Even-Numbered Five-Minute Samples

Category of Behavior	r P.E.$_r$	Category of Behavior	r P.E.$_r$
Total Job Activity	.72±.08	Total Non-Job Activity	.72±.08
Job Material	.75±.07	Non-Job Material	.51±.12
Job Self	.85±.05	Non-Job Self	.69±.09
Total Language	.74±.08	Total Physical Contacts	.78±.07
Job Language	.83±.05	Job Contacts	.88±.04
Non-Job Language	.81±.06	Non-job Contacts	.54±.12
Indeterminate Language	.54±.12	Material Contacts	.82±.06
		Non-Material Contacts	.54±.12
Social Language	.83±.05	Subject Contacts	.84±.05
Non-Social Language	.35±.15	Object Contacts	.28±.16
		Subject-Object Contacts	.86±.04

.54 ± .12. The data for contacts involving material show greater self consistency than the data for non-material contacts, the coefficients being .82 ± .06 and .54 ± .12, respectively. The correlation coefficients for language relevant to the job and for non-job language are highly similar, .83 ± .05 and .81 ± .06, respectively. Language not addressed to other persons and contacts received from persons are the categories of behavior which show least consistency, the coefficients being .35 ± .15 and .28 ± .16, respectively. The low correlation for odd and even samples of the data on object contacts is in accord with the results of previous studies and is probably due partly to the fact that observers tend to record more accurately contacts initiated by the observed child than those received by him (since the observed child was the focus of attention in all other aspects of behavior recorded) and partly to inadequate sampling. The low coefficient for non-social vocalization is perhaps due in part to the fact that this category included a number of dissimilar forms of vocalization—self speech, singing, whistling, humming, etc.—which, considered separately, might have shown greater consistency. On the whole, the kindergarten data for odd and even five-minute samples show a considerably higher degree of consistency than the corresponding nursery school data.[1]

We have seen that, in the kindergarten work period, despite the partial predetermination of behavior by the requirements of

[1] See p. 40, *supra*.

the situation, it is possible to define behavior patterns of individuals with a fair degree of assurance. Comparison of the data for this small group of children with comparable data for a somewhat larger group of the same age range will be possible in a later volume and should show more conclusively the extent to which the categories we have chosen are capable of discriminating among individuals.

CHAPTER III

A PRELIMINARY REPORT OF A STUDY OF SOCIAL-MATERIAL-SELF ACTIVITY PATTERNS IN A TRADE SCHOOL GROUP[1]

As we go from the kindergarten level to higher chronological levels it becomes increasingly difficult to find opportunities for observing spontaneous responses of individuals to their environment. Recreation groups large enough, of sufficiently long duration and with a sufficiently constant personnel to suit our purposes are rare. In the public schools, as we go from the lower to the higher grades, we find a progressively larger part of the daily program occupied by structuralized work activity, with free activity restricted to brief and infrequent recess periods.

An exception to the usual high degree of structuralization of behavior in public schools has been found in the auto mechanics class of a public trade school. The twenty boys composing this class spend their time, with the exception of one hour a day of academic work and two hours a week of electrical instruction, in a work shop which, to all outward appearances, resembles an ordinary garage. The boys are left to their own resources most of the time, receiving individual instruction when a particular job demands it, otherwise learning by doing jobs themselves or by watching others work. Constant work is not required. The boys are free to walk about the garage, to assist others, to do jobs that have not been definitely assigned, to buy candy when the boy comes to sell it, and to eat lunch when they please. There is freedom of conversation at practically all times and there are no restrictions on whistling or singing. The formation and duration of groups and the type of activity are to some

[1] The study reported in this chapter is being made by Bertha T. Hirshstein, Sterling Fellow at Yale University. It is an experimental project still in the early stages of development and is reported here because it suggests a potentially significant approach to the study of certain groups of individuals at an age level intermediate between the preschool and the adult periods. We recognize that some of the definitions of categories do not adequately meet the standards set in other studies and that the problem of proper sampling of behavior in this situation has not yet been satisfactorily solved.

extent predetermined by the fact that specific jobs are assigned to individual boys. A fact which makes this situation particularly advantageous for purposes of observation is that the boys are accustomed to being watched by the owners of cars brought to the garage for repairs. The introduction of a constant observer does not, therefore, alter the situation as much as might otherwise be expected. Although the requirements of the work situation partially restrict spontaneity of behavior in this trade school group, the possibility of free social interaction is unusually great as compared with that found in the class rooms of the average high school.

In studying the behavior of these adolescent boys a different method of approach from that used in the studies of young children has been selected, in that attention is focused upon the group rather than upon the individual. This change in emphasis has necessitated a corresponding change in recording technique. Instead of observing the same individual continuously for a five-minute period, we observe each member of a particular working group once every ten seconds during a fifteen-minute period. This method affords a sampling not only of group behavior but also of the behavior of individuals within the group.

We hope through this emphasis on groups to determine to some extent the usefulness of observational techniques in the study of groups as such. We are interested in studying work patterns and the social interaction accompanying work activity both in individuals and in groups of various sizes. Some individuals may be consistent in method of work and also in the amount of social contact accompanying their work whereas others may be consistent only in some of their activity patterns. Other individuals may be characterized by their inconsistency. Amount of social contact and frequency of involvement in groups may give some indication of "introvert" or "extrovert" tendencies. The data may also afford clues as to "dominant" and "submissive" personality tendencies. The fact that the size of group varies affords an opportunity for studying possible relationships between size and composition of groups as determinants of activity patterns and of social interaction. The number and identity of persons in the group, the constancy of

groups (within a given time) and the types of activity associated with groups of various sizes may also be studied.[1]

We have noted a change in the general approach of this study as compared with the studies described in the preceding chapters, —namely, the substitution of repeated instantaneous observations of individuals within groups for continuous five-minute observations of individuals not selected by the criterion of involvement in a group. The categories of activity to be observed have also been modified. In this situation, where the emphasis on work activity is considerably greater than in the kindergarten work period, four categories of job activity have been differentiated,—*Independent Activity*, *Coöperative Activity*, *Observation of Persons* and *Observation of Material*,—and all varieties of non-job activity have been grouped together in one category called *Extraneous Activity*.

Behavior is classified as *Independent Activity* and indicated by the symbol I, when a boy works independently, using his own tools or hands and receiving no assistance from others. Communication with other persons during work does not change the classification. Walking is recorded in this category when it is evident that the boy is going to get material related to his job and when he remains near enough to the car to be considered part of the spatial group and to permit the observation to continue. Such activities as wiping the hands or tools, opening the door of the garage (as a part of the job), working independently on one part of a mechanical device such as an auto lift while others work independently on other parts, are classified as *Independent Activity*. This category takes precedence over all others. If, for example, a boy is manipulating a tool while watching another boy, the symbol is I.

Coöperative Activity, represented by the symbol C, includes all instances of work activity where two or more boys help one another. Such coöperative activity is, by definition, simultaneous. No attempt is made to differentiate between active and passive coöperation. The behavior of a boy who holds a light

[1] That the size of groups and the types of activity observed are influenced by numerous factors other than personality differences is recognized. The number of cars available to be worked on, the specific job in hand and the assignment given by the instructor are some of the factors which we know are important. Our assumption is that, regardless of these factors, personality differences will be shown by the records, if behavior is sampled over a long enough period of time.

while others work is classified in the same way as that of the boys who are functionally manipulating materials. The following types of activity are classified in this category,—holding a fender while another boy hammers it, exchanging tools, watching head lights or stop light while another boy manipulates the electrical system of the car, adjusting brakes while another boy puts them on or off, opening the garage door so that a car can be driven out. At times verbal or other overt evidence of coöperation is necessary. If one boy starts a car while another is watching the engine, it can not be assumed that the two are necessarily coöperating unless there has been some overt communication to indicate that both are concerned with the same part of the mechanism. Mere shifting of attention, as indicated by the direction in which the eyes are focused may cause a difference in classification. If two boys are testing a radiator for leaks, one watching for leaks, the other, pouring water, the symbol is C for both. If, however, the boy who is watching shifts his attention to the boy pouring the water, his activity is classified as *Observation of Persons*, and the classification for the other boy becomes *Independent Activity*.

Definite attention to the job is a prerequisite for classification in the category *Observation of Persons*. The symbol O is recorded when a boy is watching other members of his own group work. Watching another group is considered extraneous activity.

Observation of Material, represented by the symbol Ø, includes all observation of job materials not being used by persons. When a boy is conversing about the job without working or observing others work, his behavior is arbitrarily recorded in this category whether he is looking at material or not.

Behavior is recorded as *Extraneous Activity*, with the symbol E, whenever a boy is distracted from the particular job in hand. The following types of activity are considered extraneous,—conversation about matters irrelevant to the job while not observing job materials or the members of the group at work, eating and not observing materials or persons at work, staring into space, watching a group other than the one to which the boy belongs while still remaining in the same spatial group. When a boy walks away from the car on which he has been working to a place where he can not be observed, the behavior is arbitrarily

classified as *Extraneous Activity*, whether there is evidence that the activity is related to the job or not. The record for any member of a group is discontinued as soon as he leaves the group and is not started again until he returns, so that the arbitrary decision usually affects only one interval. If extraneous activity occurs simultaneously with job activity (e. g., eating a sandwich and observing job material), the appropriate symbol for the job category (I, C, O or Ø) is recorded rather than the symbol for extraneous activity.

Language is recorded when it occurs, in addition to the symbol classifying the behavior in one of the five categories which have been enumerated. A line in the upper left hand corner of the recording space indicates that the observed boy is the speaker and a line in the lower right hand corner indicates that another person has addressed him. The record of language is in terms of occurrence only, not of content, but the content of the language sometimes helps to determine the category in which behavior is classified.

A group to be studied at any given time is defined spatially to include all boys working at or observing a given car or any one standing within this spatial area. When two cars are so close together that it is impossible to tell to which group an individual belongs, the decision is made on a functional basis if possible, i. e., the individual belongs to the group with which he is working or which he is observing. A boy working with one group and talking with another, or vice versa, is considered a member of the group being observed regardless of which one it is, if the groups are close enough together to be considered one spatial group. A person is considered a member of a group if he stops to observe the group or talks with a member of the group. A person who merely passes by is not considered a part of the group. There must be a pause, handing of a tool, communication, or other evidence that the person has temporarily become a part of the group.

An instantaneous observation is made of each boy in the group once every ten seconds. When there are only two boys in the group, the rule is to observe each one at the beginning of the period and then to look at the watch rather than at the boys for the rest of the period. This is a precaution against recording

additional conversation that does not occur at the time of the original observation. The members of the group are observed in a definite order, that of the arrangement of initials on the record form. This assures a regularity of observation within the group. Groups are also rotated as far as possible. If one is observed twice during a morning, the others are also observed twice. Since groups are not constant for any period of time in this situation, an absolutely consistent scheme of rotation can not be maintained. The number and variety of cars are constantly changing as new jobs are brought in and old ones completed. When all members of the observed group leave the car on which they have been working, a period of two minutes is allowed for the group to return. If they have not returned by the end of that time, the record is discontinued. If, before the end of the two minutes, it becomes apparent that the group has left permanently, the record is stopped. When only one boy remains, his record is continued until the others return, if within the two-minute limit, or until it is apparent that they are not going to return. If a boy is under or inside a car and can not be seen, it is assumed that he is working, if there is language, movement, or other evidence, but this assumption is recognized on the record by the symbol ⊓ indicating the beginning and the symbol ⊔ indicating the end of the time when the boy could not be seen.

The record form contains a separate column for each boy and one for the instructor, headed with their initials, and additional blank columns for other persons who may enter the group during the observation period. The sheet is divided horizontally into five sections, each representing one minute. These in turn are divided into six ten-second intervals, as indicated by the time column at the left of the sheet. One of the five category symbols is recorded in each ten-second interval for each boy in the group. If the category remains the same in successive intervals, a check is used in place of the symbol.

Chart VIII represents a sample record for the first five minutes of a fifteen-minute observation period, beginning at 9:15 A.M. on March 14, 1933. The job activity was polishing a Buick car and the group was composed of three boys, JF, RM and ST. A fourth boy, DD, joined the group three times during this period, remaining in the group not more than thirty seconds at

CHART VIII. SAMPLE RECORD OF THE JOB AND NON-JOB ACTIVITIES OF THE BOYS JF, RM, ST, AND DD FOR THE FIRST FIVE MINUTES OF A FIFTEEN-MINUTE PERIOD

DATE 3/14/33 TIME 9:15 JOB POLISHING BUICK PAGE NO. 1

TIME	JA	NB	RC	SC	DD	NF	JFr	JF	HG	PG	JH	ML	JL	RM	LM	VR	ST	FT	EU	G	Pat						
1-1-0								C						I			I										
2 - 10								I						✓			✓										
3 - 20								✓						✓			✓										
4 - 30								✓						✓			✓										
5 - 40								✓						✓			✓										
6 - 50								✓						✓			✓										
	JA	NB	RC	SC	DD	NF	JFr	JF	HG	PG	JH	ML	JL	RM	LM	VR	ST	FT	EU	G	Pat						
7-2-0					O			✓						✓			✓										
8 - 10					✓			φ						✓			✓										
9 - 20					✓			/O						✓			✓										
10 - 30								I						✓			/φ										
11 - 40								✓						✓			O										
12 - 50								✓						✓			/✓										
	JA	NB	RC	SC	DD	NF	JFr	JF	HG	PG	JH	ML	JL	RM	LM	VR	ST	FT	EU	G	Pat						
13-3-0								✓						✓			✓										
14 - 10								✓						✓			φ										
15 - 20								✓						✓			✓										
16 - 30								/✓						/C			I										
17 - 40								φ						I			✓										
18 - 50					O			/✓						✓			✓										
	JA	NB	RC	SC	DD	NF	JFr	JF	HG	PG	JH	ML	JL	RM	LM	VR	ST	FT	EU	G	Pat						
19-4-0					E			I						✓			✓										
20 - 10								E						✓			✓										
21 - 20								✓						✓			✓										
22 - 30								✓						/✓			✓										
23 - 40					I			I						✓			✓										
24 - 50					O			✓						✓			✓										
	JA	NB	RC	SC	DD	NF	JFr	JF	HG	PG	JH	ML	JL	RM	LM	VR	ST	FT	EU	G	Pat						
25-5-0								E						✓			E										
26 - 10								✓						✓			✓										
27 - 20								✓						✓			I										
28 - 30								C						O			✓										
29 - 40								I						✓			✓										
30 - 50								✓						φ			✓										

any time. DD did independent work only once during this time (Interval 23). The rest of the time he was observing the others or was in the process of leaving the group. He has no record of verbal communication with the group. The three boys present at the beginning of the observation remained the entire time, each working independently for the most part. Three instances of coöperation were noted, two for JF (Intervals 1 and 28) and one for RM (Interval 16). In Interval 20, JF was distracted from job activity and in Interval 25 both he and ST were distracted. RM's record shows continuous concentration on job activity. JF has the largest number of changes in type of activity. Language was recorded for JF in three intervals, for each of the other boys, in two intervals.

At times, because of the complexity of the observed situation, due to the number of boys in the group, to difficulty in seeing behavior or to frequent changes in type of behavior, the usual ten-second interval had to be extended to twenty or thirty seconds. The comparability of these records in which the intervals were of irregular length with those in which the regular ten-second interval was maintained is being investigated.

Whether the group record gives a representative picture of the activity of the individual is another point to be investigated. Do these records which sample behavior intermittently at ten-second intervals during a fifteen-minute period give as adequate a picture of individual behavior as five-minute records of consecutive activity? A comparable number of individual records of the consecutive type is being made, in order that the two types of record may be compared. If the data derived from the group record prove to be as reliable as the data from the individual records, the group record will be considered the better one, since more data can be obtained in the same amount of time. It is possible, however, that it will be necessary for the group records to cover a longer period of time than the individual ones, if they are to be as reliable.

CHAPTER IV

SOCIAL-MATERIAL-SELF ACTIVITY PATTERNS IN AN ADULT INDUSTRIAL GROUP

The direct observation of industrial employees during work, to be described in this chapter, in contradistinction to observation of nursery school children during free play, of kindergarten children during the work period and of trade school boys in an auto mechanics class, is set in the highly structuralized situation represented by a well organized garment-factory.[1] From the standpoint of the intense regimentation found in some factories, the complete freedom of speech, the use of the individual machine, and the lack of an automatic conveyor system, which characterized this particular factory, may suggest that the degree of structuralization was very moderate. Yet, in contrast to other situations in which we have made observational studies, it obviously imposes very definite limitations upon the expression of the worker's personality. The behavior of a worker at any given moment must, therefore, be conceived as the resultant of the structuralized work situation as well as of a large number of uncontrolled variables.

We have attempted to develop an observational technique which would produce data indicating the expression and the interplay of personalities in the structuralized work situation in terms of individual differences, which would define group differences in terms of work situations, and which would lead eventually to a correlation of these data with objective records of such items as age, mobility, progress, output and promotions.

To fulfill these aims, we sought data that were entirely in terms of overt behavior, quantitative, and reliable within known limits. The fundamental points of differentiation among individuals and the reliability of the quantitative data by means of which individuals are to be differentiated must be established before we can proceed further. Our work to date, therefore, has been limited to these two points and to the related methodo-

[1] It seems inadvisable to describe the factory in greater detail because of possible identification.

logical problems. The factory situation has proved particularly interesting methodologically in that it throws into focus some of our observational problems. Its high degree of structuralization and resultant predetermination of behavior should show to what degree our present observational technique is useful and valid under such conditions.

It was possible to apply, with necessary modifications, the same type of method as that developed in the observation of young children, i. e., a timing technique the units of which permit a quantitative differentiation in terms of the individual's concern with the material, with the self, and with the person components of his environment. A preliminary study was made in 1929–1930, in which two small groups of workers in the kitchen of a large cafeteria were observed daily for a period of eight months.[1] Subsequently, observations were made of a larger number of workers in the above-mentioned factory for two hours daily over a period of six months.

In studying adults it is of fundamental importance to know how far the observer's presence modifies the situation and biases the records obtained. If it changes the situation, either by stimulating workers to exhibitionistic performance, i. e., playing to the grandstand, or by imposing restraint through the suggestion of spy service, the fundamental objective of obtaining information about "normal" social interaction and "normal" work activity is defeated. Whereas young children can be conditioned to apparent indifference by the inconspicuousness of skillful observers, the adult's reaction to observation must be an active acceptance, if the observer is to succeed. There are undoubtedly limitations, varying with the individual and the situation, beyond which this acceptance will not go. Rejecting subterfuges of concealment as unworthy and entirely unreliable,[2]

[1] One of these groups was later used for testing the reliability of observers and of the technique.

[2] While our work is not a "time study" as known in industry, the acceptance by the worker is equally necessary. "Perhaps it should be recalled right at the start that effective time study has never been conducted outside the worker's knowledge and participation, and that the coöperation of the particular individual whose work is being analyzed is requisite to satisfactory results. Now, however, it is becoming evident that group reactions vitally condition the effectiveness of time study. More and more we recognize time study as one factor in a total situation and not as a thing apart. Morale—the state of mind of the group—has become precedent to the proper practice of time study and to any reasonably accurate determination of production standards for the individual."—Cooke, Morris Llewellyn, "Morale as a Factor in Time Study Technique," Bul. of the Taylor Society. 1927. Apr. XII, 2, p. 331.

we have adopted a policy of frankness in regard to the stop-watch and have not discussed our results even in their most technical aspects with anyone concerned in the situation. The coöperation of the chief executive, who approved of an observational program on a strictly experimental basis and expected no practical results for his own use, and the personal confidence of the minor executives were absolutely essential to securing rapport. The period of adjustment was seemingly brief, but certain changes in the data, and confidential statements from exeutives most closely in touch with the employees, led us to realize that the acceptance of an observer with a stop-watch as a harmless individual had taken longer than we had believed.

The two main categories of adult behavior in industrial situations are obviously job activities and non-job activities. Within each of these categories occur those activities which are important for our purpose, i. e., which will presumably be significant for the differentiation of social-material-self interests. Each of the major categories was, therefore, sub-divided into three functional parts representing behavior connected with materials, with the self, and with other persons (i. e., into the categories—*Job Material, Job Self, Job Person, Non-Job Material, Non-Job Self* and *Non-Job Person*). To insure a continuous and accurate record, we were forced to add a third major category to include all behavior which the observer, because of defects in the immediate observational situation or peculiarities of the worker, could not identify as either related or unrelated to the job. This category was labeled *Indeterminate.* In the *Language* category speech addressed to other persons was recorded whenever it occurred, together with the identity of the person addressed, but with no reference to the content of the speech. The recording of non-social vocalization, such as humming, was eliminated because the preliminary study had established its unreliability.

So far, except for the addition of the *Indeterminate* category and the elimination of non-social vocalization, the technique is markedly similar to that used in the kindergarten study described in Chapter II. Since, however, the broad categories thus far mentioned did not differentiate individuals sufficiently in the highly structuralized factory situation, three potentially significant sub-categories were added. The occasional simultaneous

occurrence of job and non-job behavior (e. g., a worker kept her machine going while she arranged her eyeglasses) suggested the addition of the category *Simultaneous Job and Non-Job.* A category called *Active Self* differentiated such behavior as rubbing of eyes, arms, head, etc. from passive self behavior (daydreaming, doing nothing). Observation of other persons was considered as important as speech and was recorded whenever it occurred. This behavior, in the later analysis, is labeled *Observation of Associates.*

The specific limitations of the job categories were defined for the particular factory situation on the basis of a knowledge of current practice. The essential points in these definitions, however, were in such terms as to be generally applicable to other work situations. Job activity included any overt behavior leading to the production of a specified product through change in the form or position of material, as, for example, operation of a machine, going for supplies or asking for assistance in work. Non-job activity included all other behavior.

The distinction between the *Material* and *Self* categories was kept in the same terms as that accepted in the other situations we have studied. Material activity included all functional use of materials which involved movement of the object (manipulation and transportation) but excluded support of, or by, material (e. g., pointing with a pencil or sitting in a chair). Self activity included non-functional use of materials and support of, or by, material objects, as well as all behavior concerned neither with materials nor with persons. Personal contact through material (e. g., the transfer of an object from one person to another) was recorded under both *Material* and *Person.* Other personal contacts (e. g., shaking hands) were recorded only under *Person.* The *Job* and *Non-Job* distinction was maintained throughout and the rôle of the person observed (subject or object) was indicated when personal contacts occurred.

Job materials included all materials used in the production of the job end product. Non-job materials comprised all others. The immediate use of the materials determined the classification. For marginal situations, arbitrary rules as reasonable as possible were made and consistently applied. Familiarity with the requirements and practice in the work place was necessary in mak-

ing these finer distinctions. Making a record on a work coupon, cutting it off and placing it in a box, for example, was a series of activities recorded under *Job Material*. Taking coupons out of the box and counting them were activities which were recorded under *Non-Job Material*, since they were not a part of the job in this factory, and, in the case of most workers, represented a personal activity occurring during recess when work was forbidden.

The two extremes in job and non-job activities were easily distinguishable,—(1) job activity—working silently with materials or walking without distraction to get materials, and (2) non-job activity—doing nothing, no apparent preoccupation with job materials or with people. The marginal stages between job and non-job behavior often involved the use of speech or of observation (either of work or of associates). The type of behavior concurrent with self activity determined the classification as *Job* or *Non-Job*.

Occurrence of language was recorded in the *Material* or the *Self* column in the appropriate five-second interval. The distinction between *Job* and *Non-Job* which had been applied in the kindergarten was not attempted because of the impossibility of hearing what was said clearly enough to permit such a differentiation. Thus, if the worker was in functional contact with job materials and was at the same time talking to other workers, the number identifying the person addressed was recorded in the *Job Material* column, irrespective of whether the content of the speech pertained to personal or to job situations. Speech and gesture, however, were often determining factors in the placement of behavior in the *Job Self*, *Non-Job Self* or *Indeterminate* categories. If the worker, when not using job materials, called the supervisor, or beckoned to him, the main behavior classification was *Job Self*, and the symbol for language (the individual's identifying number) or for observation (g) was added. If the speech was inaudible, was not accompanied by gesture toward a supervisor, and resulted in no observable response in any way related to the job, the behavior was recorded as *Indeterminate*, with the language symbol. If the speech was clearly understood as personal, the behavior was recorded under *Non-Job Self*, with the language symbol. It should also be noted that, when there

was no functional contact with materials, listening to the speech of supervisors determined the classification of the activity as *Job Self*. Listening to an associate talking about his work or personal affairs was, of course, classified as *Non-Job Self*.

Observation of work or of associates constituted one criterion for classifying a specific activity as *Job Self, Non-Job Self*, or *Indeterminate*. Watching demonstration of work or examining job materials, when there was no functional contact with materials, was classified as *Job Self*. Looking around for a person or object was classified as *Job Self*, if immediately followed by speech, gesture, or physical activity concerned with the job; as *Non-Job Self*, if closely associated behavior related the activity to personal interests; and as *Indeterminate*, if there was no clear-cut basis for classification in work or personal categories. Observation of work which was accompanied by functional use of materials was not distinguished from other forms of job material activity.

Simultaneous job and non-job activities occurred in the following situations,—walking in the course of work, listening to directions or asking questions regarding work while manipulating non-job materials (e. g., adjusting clothing). This type of behavior was classified as concurrent *Job Self* and *Non-Job Material*. Simultaneous use of job and non-job materials (e. g., operating a machine and adjusting eyeglasses) was classified as concurrent *Job Material* and *Non-Job Material*. Functional contact with job materials while rubbing the eyes or arranging hair was classified as *Job Material* and *Active Self*. Job self and active self activities might also occur concurrently and were distinguished in the record.

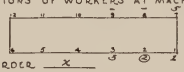
We attempted to equalize the effect upon workers of as many as possible of the variables which would interfere with the differentiation of individuals in the terms in which we were interested. By selecting for observation all the regular machine workers in three work centers having somewhat similar machines and belonging to one department, we entirely eliminated one variable, sex, and decreased those of supervision, kind of work, and work associates. Yet sub-foremen and inspectors exerted irregular influences, workers did varying amounts of non-machine work, and work associates were far from constant. A minimum opportunity for social interaction was assured by observing no worker

unless there was at least one other regular worker at an adjacent machine. At the beginning of each observation the position of the worker in relation to others and the number of her machine were recorded, and any change of positions during the observation period was noted. Availability of work, as indicated by work material at hand, was a requisite for continuation of the record. Thus, if the supply of job material before a given worker was exhausted temporarily, even for two intervals, i. e., ten seconds, the observation was interrupted. But if the worker turned at once to call for a supply, and, while waiting for it, did not enter into a conversation about personal matters or other non-job activity, the classification was *Job Self*. The effects of work fatigue and of end spurts, the latter said to be very pronounced in the late afternoon, were avoided by observing only up to the first rest period in the middle of the forenoon. Observations were confined to the time beginning fifteen minutes after work started in the morning, which was 7:30 A. M., and ending five minutes before the preliminary bell for the rest period. A mechanical order of rotation for both individuals and work groups was followed. We also tried to equalize such variables as approaching holidays, varying pressure of work, the weather, minor accidents, and changes in local personnel by distributing our observations over a period of several months.

Arbitrary rules of recording overcame some factors which would otherwise have skewed the record. For instance, when workers left the room, the non-job self record was uniformly continued for the same length of time before the record was broken, regardless of the section of the floor from which workers came. When a repair man interrupted job material activity for more than thirty seconds the observation was arbitrarily broken. Interruptions of this kind rarely exceeded twenty or thirty seconds, and, when they did, the effects were varied. One worker would wait quietly, another would visit and a third would go to another machine. The influence of a particular condition can be measured, of course, only if it has been recorded every time it has been present. The only conditions we have recorded consistently enough to use in analysis of the present data are the number of the machine and the members of the immediate group.

DATE 3-24-32 TIME 10.25

POSITIONS OF WORKERS AT MACHINES



RECORDER x

TIME	NON JOB			JOB		
	PERSON	SELF	MAT	MAT	SELF	PERSON
1-1-0						
2 - 5						
3 - 10						
4 - 15						
5 - 20					1	
6 - 25						
7 - 30						
[illegible] - 35						
9 - 40			1		3	
10 - 45					3	
11 - 50						
12 - 55						
TIME	PERSON	SELF	MAT	MAT	SELF	PERSON
13-2-0						
14 - 5					g	
15 - 10						
16 - 15					π	
17 - 20					π	
18 - 25					π	
19 - 30						
20 - 35						
21 - 40						
22 - 45						
23 - 50			1			
24 - 55						
TIME	PERSON	SELF	MAT	MAT	SELF	PERSON
25-3-0						
26 - 5						
27 - 10						
28 - 15						
29 - 20						
30 - 25						
31 - 30						
32 - 35					g	
33 - 40						
34 - 45						
35 - 50						
36 - 55						
TIME	PERSON	SELF	MAT	MAT	SELF	PERSON
37-4-0						
38 - 5						
39 - 10						
40 - 15					3	
41 - 20					3	3
42 - 25						
43 - 30						
44 - 35		g				
45 - 40		3				
46 - 45		3				
47 - 50		g				
48 - 55						
TIME	PERSON	SELF	MAT	MAT	SELF	PERSON
49-5-0						
50 - 5						
51 - 10					g	
52 - 15						
53 - 20						
54 - 25						
55 - 30						
56 - 35					1	
57 - 40						
58 - 45						
59 - 50						
60 - 55						

CHART IX. SAMPLE FIVE-MINUTE RECORD OF THE JOB AND NON-JOB ACTIVITIES OF WORKER 2

Chart IX represents the record of Worker 2 made on March 24, 1932 by Recorder X. The record was begun at 10:25 with Worker 2 seated at Machine 2 and Workers 1, 3 and 5 at Machines 1, 3 and 7 respectively. No change in positions occurred during the observation. Worker 2, with one interruption in Interval 5, worked continuously with job materials from the beginning of the record until the end of Interval 15. During a part of Interval 5 she carried on job activity without functional use of materials. In Interval 9, concurrently with job material activity, she used non-job materials (as indicated by the parallel lines in the *Job Material* and *Non-Job Material* columns) and spoke to Worker 3, continuing the speech in the following five-second interval. In Interval 14 (Minute 2), she glanced at her associates (g). At the beginning of Interval 16, she ceased to use materials and spoke to the group (π), continuing the speech in the next two intervals. Since there was no means of identifying her behavior in the next two and a fraction intervals, a line was drawn in the *Person* column to indicate that the behavior was indeterminate.[1] In Interval 23 she ceased job activity for a few seconds to use non-job material. The only deviation from continuous job material activity in Minute 3 was a glance at other workers in Interval 32. In Intervals 40 and 41 (Minute 4) Worker 2 spoke to Worker 3 and in the latter time interval there was an exchange of job materials initiated simultaneously by the two workers.[2] In Intervals 44–47 of the same minute, a period of slightly more than fifteen seconds, she ceased job activity completely, spoke to Worker 3 in two intervals and in two other intervals looked at other workers without speaking. In the last minute job activity was continuous. Worker 2 glanced at other workers in Interval 51 and carried on work without use of materials in Interval 56. The only type of activity not represented in this record is active self behavior. This activity was recorded by an irregular line in the *Non-Job Self* column.

We first made 16 five-minute observations on 20 workers

[1] To avoid the addition of an extra column for indeterminate behavior, this device was used. A continuous line in the *Person* column, unaccompanied by any symbol representing a person, always meant indeterminate behavior. Contacts with persons were shown by the identifying number of the individual and the appropriate symbol indicating the type of contact, as in Interval 41.

[2] The same symbols were used for the subject, object, and subject-object types of contact as in the kindergarten study. See p. 54, *supra*.

within a period of two months. This group of observations constitutes Sample I. Sample II consists of a similar number of observations on the same 20 workers made during the next two months. Sample III consists of observations on the same workers for two forty-minute periods each, made during the course of the following two months with the purpose of obtaining a picture of consecutive activity. Of these three samples, the first and second were made under as nearly as possible the same conditions. The third sample differs from the others in the following respects. It was collected in a period in which there were at times unusually large orders of work. The two forty-minute observations were made on two days instead of being distributed over 16 days. One difference was made in the conditions of observation, namely, the absence of an adjacent worker did not lead to an interruption of observation, although the record was never begun unless another worker was adjacent. The data from the first and second samples, collected under similar conditions, are combined in all analyses except the preliminary analysis of Sample I and the comparison of samples.

To summarize, the types of data available from these records were in terms of the time-unit frequencies of each of six main categories,—*Job Material*, *Job Self*, *Job Person*, *Non-Job Material*, *Non-Job Self* and *Non-Job Person*. The *Non-Job Self* category was further refined to differentiate between active and inactive behavior. Behavior that could not be identified as either *Job* or *Non-Job* was classified as *Indeterminate*. Language and observation of associates were recorded in the appropriate time interval. When a preliminary analysis of Sample I was made, it was found useless to proceed further with the analysis of the data classed as *Person* (i. e., physical contacts) since these had occurred in only 37 of 20,000 units of behavior. Furthermore, they were almost invariably necessitated by a particular job situation, 9 of the 37 being made by one individual whose work occasionally required the transfer of objects. We accepted this fact as confirmation of our earlier tentative conclusion that, in these industrial situations, the physical contact of individuals is unimportant because of its infrequency and also because of its dependence on the job.

While 1.5% of the total job activity recorded in Sample I did

not require the active manipulation of materials, the occurrence of such job self behavior was very irregular and was apparently determined by the job. A worker might in half a dozen consecutive records use job materials without a break and then in the next record she might spend as much as 10% of her working time without any functional contact with materials. This apparent influence of the job rather than of the worker in determining the division of job activity into material and self components led us, in studying individual differences, to combine *Job Material* and *Job Self* into a single *Total Job* category.

Unlike job activities, the larger portion of non-job behavior was of the self rather than of the material type since the worker, when not carrying on the job, was rarely using materials of any kind. The occasional use of non-job materials was determined by conditions which our rotation of observations did not sample fairly and which were probably trivial from the point of view of our main interest (e. g., the worker's use of a handkerchief in the care of a cold or her use of a vanity case in preparation for the rest period). We, therefore, combined these two categories for the present analysis, thus obtaining a composite *Total Non-Job* category. We were also able to combine certain other categories which seemed to offer possibilities of characterizing individuals. We have already mentioned the recording of two simultaneous activities, one concerned with the job and the other of a personal nature, i. e., the *Simultaneous Job and Non-Job* category. Occasionally, active self behavior was carried on during work activity. This suggested a category of *Active Self Simultaneous with Job.* Since there seemed to be a fundamental relation between speech to associates and observation of associates without speech, these two categories of behavior were combined as *Contact with the Group.*

Frequency of occurrence for each of these combinations as well as for the separate categories, *Indeterminate*, *Language*, *Active Self* and *Observation of Associates*, in terms of the average number of five-second intervals per 60 intervals (five minutes of observation), was computed for each of the 20 workers. In Table VIII we present for each worker the average frequency of occurrence, in Samples I and II combined, of these nine kinds of behavior, together with the group mean and median per category

as measures of central tendency, and the group range per category as an indication of dispersion.

TABLE VIII

Average Number of Five-Second Intervals per 60 Intervals in which Different Kinds of Behavior Occurred in the Records of 20 Industrial Workers (Samples I and II Combined)

Worker	Total Job	Total Non-Job	Indeterminate	Language	Simultaneous Job and Non-Job	Active Self	Active Self Simultaneous with Job	Observation of Associates	Contact with Group
1	56.6	0.8	2.9	8.9	0.4	0.7	0.4	3.3	12.2
2	59.3	0.2	0.5	10.8	0.1	0.1	0.1	3.8	14.5
3	56.5	1.3	2.2	11.1	0.6	1.1	0.5	4.3	15.4
4	55.5	1.4	3.2	22.2	0.2	0.4	0.1	3.2	25.4
5	59.8	0.2	0.0	1.4	0.4	0.4	0.3	2.4	3.8
6	55.0	3.3	1.7	6.6	0.5	1.4	0.4	13.2	19.8
7	58.8	0.7	0.5	2.1	0.1	0.2	0.1	6.3	8.4
8	53.9	1.7	4.4	7.8	0.4	1.0	0.3	10.0	17.8
9	57.4	0.6	2.0	4.5	0.3	0.7	0.3	6.7	11.3
10	55.3	2.3	2.4	9.7	0.1	0.4	0.1	7.8	17.4
11	57.2	0.3	2.5	7.4	0.3	0.5	0.3	5.5	12.9
12	59.3	0.6	0.1	3.7	0.3	0.1	0.1	8.2	12.0
13	58.3	0.3	1.4	5.6	0.3	0.6	0.3	7.3	12.9
14	56.2	2.4	1.5	11.0	0.3	0.4	0.2	12.6	23.6
15	59.2	0.6	0.2	5.1	1.1	1.3	1.0	3.7	8.7
16	59.4	0.3	0.3	4.7	0.8	0.8	0.7	3.3	8.0
17	57.8	0.6	1.6	5.7	0.5	0.8	0.4	9.3	14.8
18	59.1	0.7	0.2	4.3	0.7	0.9	0.6	12.0	16.3
19	59.1	0.7	0.3	4.5	0.6	0.6	0.5	5.4	10.0
20	58.9	0.3	0.8	3.3	0.2	0.3	0.2	5.5	8.8
Group Mean	57.6	1.0	1.4	7.0	0.4	0.6	0.3	6.7	13.7
Group Median	58.0	0.6	1.4	5.6	0.4	0.6	0.3	5.9	12.9
Group Range	53.9–59.8	0.2–3.3	0.0–4.4	1.4–22.2	0.1–1.1	0.1–1.4	0.1–1.0	2.4–13.2	3.8–25.4

As would be expected, the great preponderance of activity is related to the job. Of the behavior recorded in the factory, 96.0% (57.6 intervals in 60) was of this type and only 1.7% (one interval in 60) could be definitely classified as Non-Job. Indeterminate[1] behavior amounted to 2.3% (1.4 intervals in 60) of the total. Indeterminate, or unidentifiable, behavior may be a characteristic of certain individuals as well as a result of conditions of observation. Nine workers had less than one interval in 60 classified as *Indeterminate*. Each of these workers had also less than one interval in 60 of non-job activity. Six workers had more than two intervals in 60 where the activity was classed

[1] In the cafeteria kitchen, where the observers were very close to the workers, indeterminate behavior was negligible in amount. Job activity included 98.9% of the total behavior recorded, non-job activity, 1.1%.

as *Indeterminate,* an average of 4.9% of their total time. The most extensive application to work was on the part of Worker 5 who spent 59.8 intervals in 60, or 99.7% of the time observed, in job activity. In two and two-thirds hours of observation distributed over four months she averaged one second per five minutes of non-job activity and never had behavior classed as *Indeterminate.* At the other extreme was Worker 8 whose average for job activity was 53.9 intervals in 60, or 89.8% of the time observed. Her activities were classed as *Indeterminate* in 4.4 intervals per 60, or 7.3% of the time. Worker 6 had the largest amount of non-job activity, 3.3 intervals in 60, or 5.5% of the time.

Simultaneous job and non-job activity, i. e., concurrent work and personal activities, averaged less than half an interval in 60 for the group, or 0.7% of the total time. A large part of this behavior was included in the category *Active Self Simultaneous with Job.* Certain workers, however, frequently adjusted eyeglasses or clothing while continuing their work, that is, they carried on concurrently non-job material and job activities. Although the amounts of behavior of this type were small, there was considerable range in the amount of total simultaneous work and personal activity,—from 0.1 interval per 60 for Workers 2, 7, and 10 to 1.1 intervals in 60 for Worker 15, that is, from 0.2% to 1.8%. These individual variations from the group average in behavior which may be relatively independent of the requirement of the job add to the significance of the relatively high self consistency in these categories, a point to which we will return later.

The group average for language was 7.0 intervals in 60 (11.7% of the total time). The range was from 1.4 to 22.2 intervals (2.3% to 37.0%). A scatter diagram and the Pearson coefficient of correlation between *Total Job* and *Language* for the 20 workers, —.58 ± .10, show that the amount of workers' conversation, for the group as a whole, was inversely related to the amount of time devoted to job activities. There were, however, conspicuous exceptions to this inverse relationship, since some individuals both worked more continuously and talked more frequently than their associates. Worker 2, for example, ranked 3.5 in application to the job and 4 in use of language.

Observation of the group had been defined in such a way as to exclude observation simultaneous with speech, so that the addi-

tion of these mutually exclusive categories to get an estimate of the extent of contact with the group was justified. While the average occurrence of these two types of activities, use of language and looking at associates without speech, was very similar (7.0 and 6.7 intervals per 60, respectively), individual workers varied as to the relationship between these two methods of involvement in the group. When the 20 workers were ranked in order of frequency of language and frequency of observation of associates, only three workers were in the same quartile for both types of behavior. Two were in the highest quartile for one behavior and the lowest for the other, and the other 15 varied by one or two quartiles. Worker 14 ranked third in frequency of use of language and second in observation of associates, having an average of 11.0 intervals of behavior per 60 (18.3%) in the first category, and 12.6 intervals (21.0%) in the second. Worker 5, on the other hand, ranked twentieth in each category, with 1.4 and 2.4 intervals of the two types of behavior per 60 (2.3% and 4.0% respectively, of the time observed). The first individual was in some kind of contact with the group nearly 40% of the time, the second, only 6%. They resembled each other only in their self consistent ranking in frequency of the two kinds of behavior. Quite the contrary behavior characterized Workers 4 and 18. Each of these expressed her interest in the group predominantly in one or the other manner. Worker 4, whose high frequency of speech has been noted, ranked nineteenth in observation of associates without speech, while Worker 18, who ranked third in frequency of observation, used speech so infrequently that she ranked sixteenth in this kind of behavior. The average number of intervals per 60 and the percentage frequency of occurrence of these two kinds of behavior for the four workers was as follows.

	Language		Observation	
	Number of Intervals	%	Number of Intervals	%
Worker 14	11.0	18.3%	12.6	21.0%
" 5	1.4	2.3%	2.4	4.0%
" 4	22.2	37.0%	3.2	5.3%
" 18	4.3	7.2%	12.0	20.0%

It is of interest to find in the adult work situation, with all its limitation of the expression of personality, the same general patterns of variation in contacts with associates which were found in the nursery school children's choice of language and physical contacts as means of communication with the group. Some individuals predominantly use one means of contact in preference to another, others predominantly select the means of contact neglected by the first group, and a third group exhibits a positive relationship between the two in the extensive or in the limited use of both avenues of contact.

The intercorrelations of different kinds of behavior are inextricably associated with self consistency and group consistency. Does a worker tend, day after day, to pattern his behavior in a particular way? In a group, what is the average occurrence of different aspects of behavior and how does a worker's deviation in one item tend to be associated with his deviation in another?

We explored, to a limited extent, the relationship between application to the job and contact with associates. As mentioned previously, the correlation between the average number of intervals of job activity and of language for the 20 workers, in Samples I and II combined, was $-.58 \pm .10$, but when we combined language with observation of associates, as a measure of group contact, the correlation with job activity became $-.74 \pm .07$. Individual variations in this relationship in each of 32 successive five-minute observations are shown for Workers 3 and 2 in Figure 4. Worker 3 varied greatly from her own mean in individual observations; Worker 2, much less. Worker 3 had an inverse relation between job activity and contact with the group in 23 of the 32 observations; Worker 2, in 14 of 32 observations.

The relation between continuity of job activity and use of language was further investigated by finding, for each of 20 workers, first the number of periods of different lengths during which there was unbroken application to the job, and then the average number of intervals of language per standard 60 intervals in periods of these lengths.[1] Three lengths of period were used,—(1) from 0 to 12 intervals, (2) from 12 to 60 intervals, and

[1] For periods lasting less than a minute, the frequency of language per "standard 60 intervals" was computed by finding the total recorded intervals of job activity, dividing this number by 60 to get the number of sixty-interval periods, and then dividing by this number the total intervals of language recorded simultaneously with job activity.

(3) at least 60 intervals, since job activity sometimes continued uninterruptedly beyond the five-minute period of observation. These data for each worker and for the group are presented in

FIGURE 4. Self Consistency of Workers 2 and 3 with respect to Application to Job and Contact with the Group, as Measured by Number of Intervals of Behavior per Observation in 32 Successive Five-Minute Observations

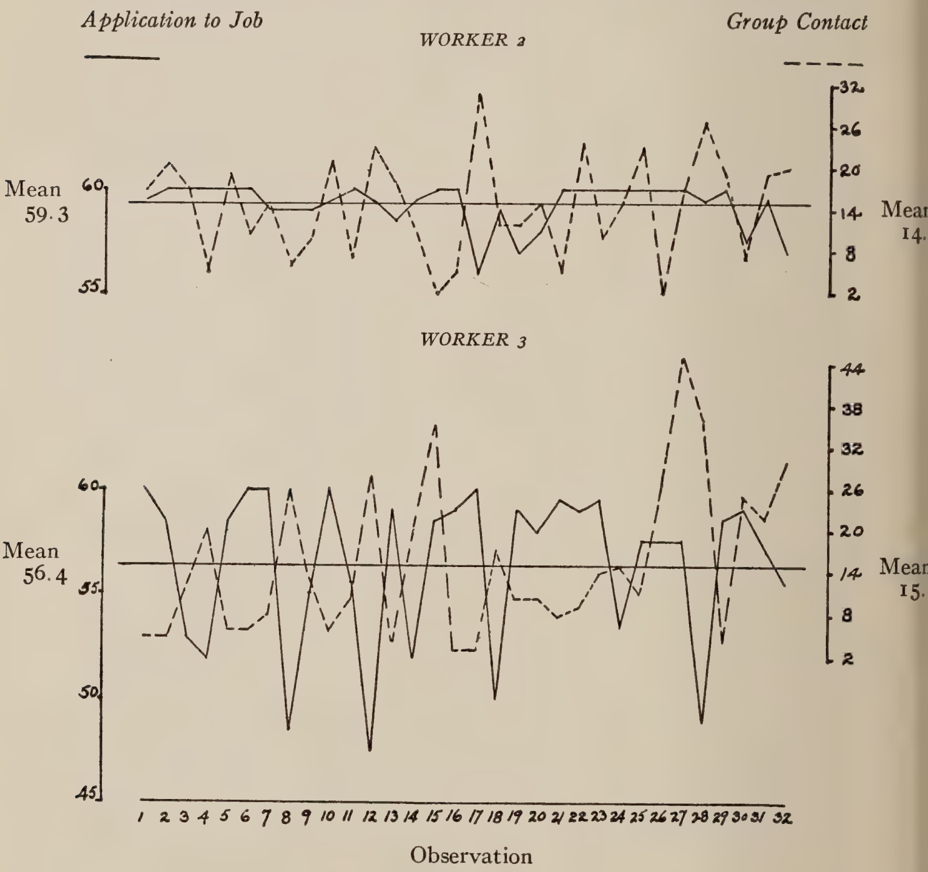


Table IX. As we would expect from the negative correlation between application to the job and use of language, the table shows a more frequent occurrence of language in the briefer than in the longer periods of continuous job activity. The group average is 2.9 intervals of language in five-minute periods of un-

TABLE IX

Number of Periods of Continuous Job Activity Distributed According to Length and Number of Five-Second Intervals of Language per Standard 60 Intervals in Work Periods of Corresponding Lengths, for 20 Industrial Workers

Worker	Number of Periods of Continuous Job Activity Lasting			Number of Intervals of Language per Standard 60 Intervals in Periods of Continuous Job Activity Lasting		
	0–12 Intervals	12–60 Intervals	at least 60 Intervals	0–12 Intervals	12–60 Intervals	at least 60 Intervals
1	49	35	6	13.8	4.6	5.7
2	22	23	16	17.3	9.1	9.6
3	68	36	5	16.1	7.4	2.6
4	68	32	8	27.0	17.8	15.4
5	7	10	27	4.2	0.6	1.3
6	55	32	9	10.3	4.4	0.8
7	19	23	18	6.4	1.8	0.8
8	66	39	6	7.0	1.7	4.8
9	41	27	14	9.6	2.4	1.0
10	54	36	6	16.9	5.7	0.7
11	41	26	14	13.4	4.9	3.8
12	13	11	24	10.6	4.1	2.8
13	24	17	18	13.0	4.5	3.1
14	49	35	7	17.9	7.6	4.1
15	28	27	16	12.1	6.3	1.6
16	20	22	18	11.9	4.9	2.8
17	41	30	14	13.7	3.8	0.9
18	20	20	21	13.5	7.6	1.8
19	19	20	19	14.5	5.0	2.3
20	19	20	19	7.3	3.9	1.1
Group Average	36	26	14	14.1	5.6	2.9

interrupted work, 5.6 intervals in periods lasting from one to five minutes (12 to 60 intervals) and in the briefest periods (less than 12 intervals) 14.1 intervals of language per standard 60 intervals of job activity. In other words, language occurred 4.8%, 9.3% and 23.5% of the time in the longest, in the intermediate, and in the shortest periods, respectively. There are interesting individual variations. Workers 2 and 10 converse with similar frequency in work periods of less than one minute in duration (17.3 and 16.9 intervals per 60, respectively) but during five-minute periods of continuous activity Worker 2 converses on an average in 9.6 intervals out of 60, or 16.0% of the time, and Worker 10 in only 0.7 intervals, or 1.2%. The latter worker withdraws from her associates, as far as speech is concerned, when she is working most continuously. Workers 6, 7 and 17 show a similar tendency to withdraw, having, respectively, 0.8, 0.8 and 0.9 intervals of language per 60 intervals of continuous job activity. When work activity is least continuous, however, Worker 6 averages 10.3 intervals of speech per 60, or 17.5% of her time; Worker 7, 6.4 intervals, or 10.7% of the time; and Worker 17, 13.7 intervals, or 22.8% of the time.

We computed also the frequency of language in periods when job activity was not recorded, i. e., in intervals of non-job and indeterminate behavior. The group average here is 32.1 intervals of language per 60 intervals, or speech 53.5% of the time. Workers 2 and 20 rank highest, with 43.0 and 42.5 intervals per 60, respectively, or slightly more than 70% of the time, but this similarity is not apparent when they are working most continuously. Worker 2 then converses on an average in 9.6 intervals in 60, 16% of the time, and Worker 20, in only 1.1 intervals, or 1.8%.

From the group tendency toward greatest frequency of speech in the shortest periods of continuous job activity, there is no deviation on the part of any worker. But, as we have noted in the case of Workers 2 and 10, workers differ in relative frequency of speech in periods of continuous preoccupation with work. Are workers consistent in their variations in frequency of language during periods of job activity of different lengths? We investigated this point by dividing each worker's observations into odd and even sub-samples,[1] and calculating for each sub-sample the frequency of speech in job activity periods of the three different lengths. A scatter diagram of frequencies of language per worker in periods of different lengths is shown in Figure 5. As would be expected, frequencies of speech for the longest periods of job activity cluster at the lower end of the distribution, and frequencies for the brief periods at the upper end. The diagram shows a considerable degree of self consistency.

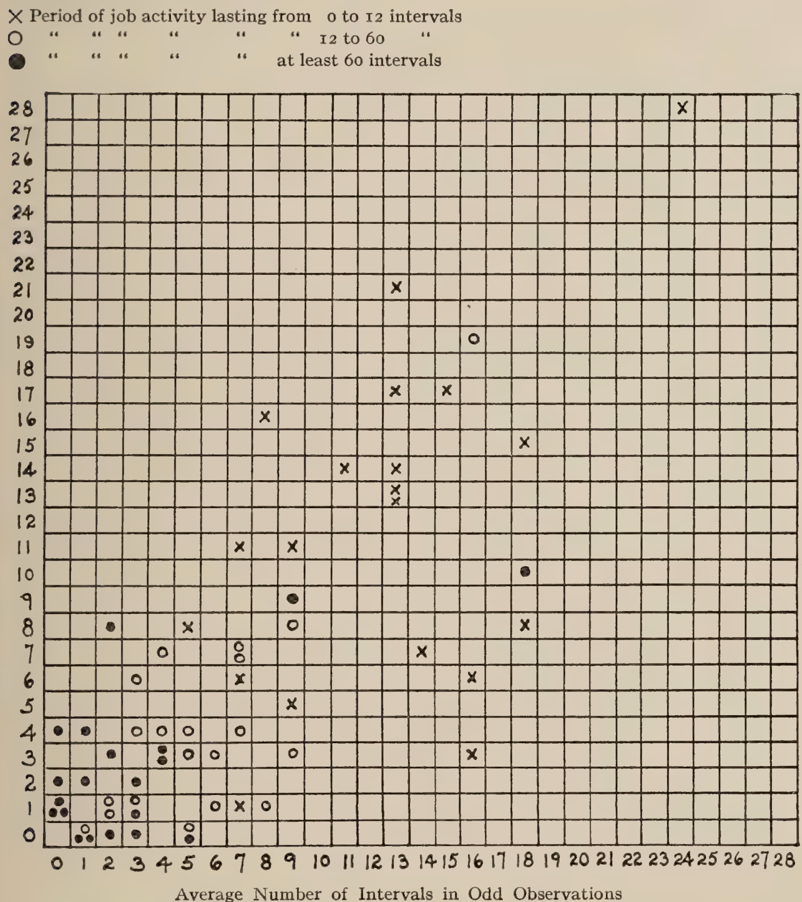
The tentative character of the categories with which we were working and the lack of adequate information as to their reliability suggested conservatism in attempting many intercorrelations of the data for different categories. The self correlation of odd and even samples of the data within the same category, collected under conditions as nearly uniform as possible, is, on the other hand, justifiable in that it throws light on the internal consistency of the data.

Although we recognized *a priori* the probable inadequacy of halves of a sample composed of only 16 five-minute observations per worker, we were interested in seeing how far consistency of behavior occurred within such samples. We were also interested

[1] Odd and even sub-samples were obtained in the same manner as that described on p. 69, *supra.*

in determining the length of sample necessary for consistency or representativeness in different aspects of behavior. We corre-

FIGURE 5. Consistency of 20 Workers in Use of Language in Periods of Unbroken Job Activity of Different Lengths, as Measured by Average Number of Intervals of Language per Standard 60 Intervals in 16 Odd and 16 Even Five-Minute Observations.



lated, therefore, the data for the nine categories of behavior, in odd and even five-minute records in Sample I and compared these correlations with those obtained when Samples I and II had been combined, i. e., we compared correlations between two

sub-samples composed of eight observations per worker (Sample I) with correlations between two sub-samples composed of 16 observations per worker (Samples I and II combined). Correlations for the two samples thus compared, with one exception, indicate highly similar degrees of consistency for the categories of activity not directly associated with the job, with a tendency toward greater consistency in the longer sample,[1] but lengthening of the sample increased to a considerable extent the self correlations in the *Total Job*, *Total Non-Job*, and *Indeterminate* categories. It is probable that, in these aspects of behavior, a sample of only eight five-minute records per worker would not differentiate workers significantly, but a sample composed of sixteen records per worker would show considerable consistency. In view of the high probable errors of the correlation coefficients, the correlations are not very meaningful. For *Total Job* the correlations for the shorter and longer samples, respectively, were .31 ± .14 and .63 ± .09; for *Total Non-Job*, .31 ± .14 and .60 ± .10; and for *Indeterminate*, .42 ± .12 and .66 ± .09. With some hesitation, we suggest that the data for both samples indicate that in this situation the use of language, the observation of associates, total contact with the group, and concurrence of personal and work activities are individual characteristics of the worker, have some degree of consistency, and are capable of repeated expression even in a well ordered structuralized situation.

One of the methodological problems in which we have been interested is the relation of data obtained by a succession of small samples to data obtained in more concentrated longer samples. In Table X we present the mean, median and range found for each category in Samples I and II (considered separately and in combination) and in Sample III. Because of the high degree of structuralization in the situation, we would expect *a priori*, that these measures would correspond closely. They show, in fact, a considerable degree of similarity. It is obvious, from comparison of the measures shown in Table X, that Sam-

[1] In Sample I, and in Samples I and II combined, respectively, the correlations were .85 ± .04 and .87 ± .04 for *Language;* .74 ± .07 and .80 ± .05 for *Simultaneous Job and Non-Job;* .84 ± .04 and .86 ± .04 for *Active Self Simultaneous with Job;* .91 ± .03 and .96 ± .01 for *Observation of Associates;* .88 ± .03 and .89 ± .03 for *Contact with the Group*. For the category *Active Self* the correlation for Sample I was .71 ± .07, for Samples I and II combined, .66 ± .09.

TABLE X

Group Mean, Median and Range per Category of Behavior for 20 Industrial Workers in Samples I, II and III, and in Samples I and II Combined

	Sample	Total Job	Total Non-Job	Indeterminate	Language	Simultaneous Job and Non-Job	Active Self	Active Self Simultaneous with Job	Observation of Associates	Contact with Group
Mean	I	57.6	1.3	1.1	6.7	0.4	0.7	0.4	6.3	13.0
	II	57.6	0.6	1.7	7.4	0.4	0.6	0.3	7.1	14.4
	III	58.5	0.7	0.9	6.1	0.4	0.4	0.3	7.0	13.1
	I & II	57.6	1.0	1.4	7.0	0.4	0.6	0.3	6.7	13.7
Median	I	58.4	1.0	0.9	6.0	0.4	0.6	0.3	5.4	12.6
	II	58.0	0.4	1.7	6.5	0.3	0.7	0.3	7.2	14.3
	III	59.1	0.4	0.4	3.0	0.4	0.5	0.3	6.1	11.8
	I & II	58.0	0.6	1.4	5.6	0.4	0.6	0.3	5.9	12.9
Range	I	54.0–60.0	0.0–4.7	0–3.6	1.1–19.7	0.0–1.4	0.0–2.0	0.0–1.3	2.2–13.1	3.3–25.8
	II	51.7–59.9	0–2.1	0–6.7	1.8–24.6	0.1–0.9	0.1–1.2	0.1–0.7	2.4–13.2	4.4–28.1
	III	54.0–59.9	0–3.5	0.0–3.3	0.9–20.1	0.0–1.2	0–1.2	0–1.0	2.2–15.1	4.2–25.8
	I & II	53.9–59.8	0.2–3.3	0.0–4.4	1.4–22.2	0.1–1.1	0.1–1.4	0.1–1.0	2.4–13.2	3.8–25.4

ple III differs very slightly, and probably insignificantly, from Samples I and II combined, and, in many cases, less than I and II differ from each other. In all three samples the median in the *Total Job* category is larger than the mean, indicating a concentration of large measures. A few workers, in fact, spent 100% of their time in work activity in as many as half of all their five-minute records. In a majority of the other categories the median was smaller than the mean, indicating a massing of the measures at the lower end of the scale. We are uncertain as to the possible significance of the slightly higher mean in *Total Job* and the slightly lower mean in *Language* in Sample III. Active self, non-job and indeterminate behavior (the two latter, obviously, in compensation for the increase in job activity) occur less frequently in Sample III, but all these differences are very slight and almost certainly within the limits of chance.

In order to compare the relative self consistency of Samples II and III, scatter diagrams were made for each of nine categories or combinations of categories. The results did not show a clear tendency for either sample to have greater self consistency in all aspects of behavior. The distributed observations (Sample II) showed greater consistency than the concentrated ones (Sample III) in some categories, less in others. The scatter diagrams for two categories illustrate this point. The observa-

tions for *Total Job* which were made on 16 different days per worker (Figure 6) show considerable variation (r = .68 ± .08), but those which were concentrated in two days per worker

FIGURES 6 AND 7. Average Number of Intervals of *Total Job Activity* per 60 Intervals in 8 and 8 Even Five-Minute Records of 20 Industrial Workers.

FIGURE 6
SAMPLE II—DISTRIBUTED OBSERVATIONS

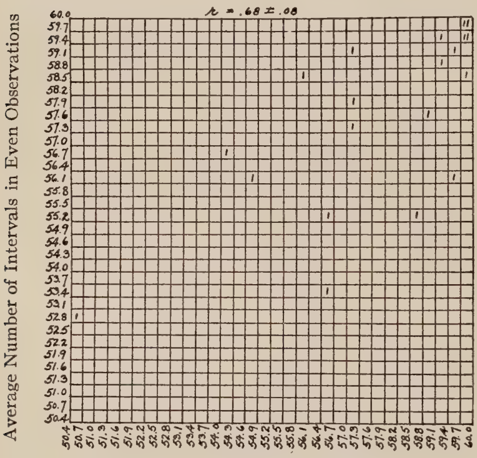
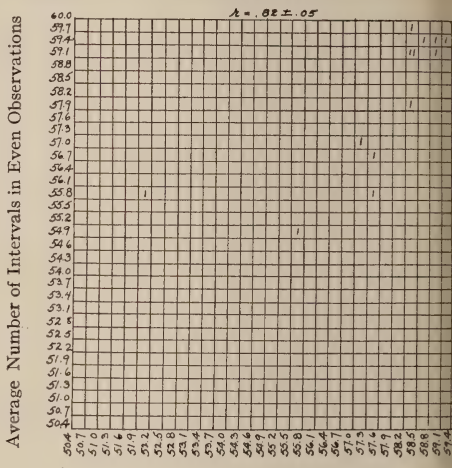


FIGURE 7
SAMPLE III—CONCENTRATED OBSERVATI



FIGURES 8 AND 9. Average Number of Intervals of *Simultaneous Job and Non-Job Behavio* 60 Intervals in 8 Odd and 8 Even Five-Minute Records of 20 Industrial Workers.

FIGURE 8
SAMPLE II—DISTRIBUTED OBSERVATIONS

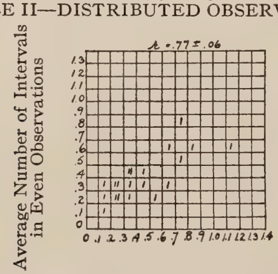
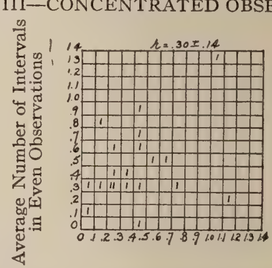


FIGURE 9
SAMPLE III—CONCENTRATED OBSERVAT



(Figure 7) show greater consistency (r = .82 ± .05). In the *Simultaneous Job and Non-Job* category we see the opposite effect of distribution (Figures 8, 9). When the data were sampled over a period of time, they had a higher consistency (r =

.77 ± .06) than when they were concentrated (r = .30 ± .14). These diagrams and correlations confirm our impression that numerous and inconstant variables affect the behavior we are recording.[1]

Observation of the social-material-self behavior patterns of forty persons in two quite different work situations does not justify extensive generalization as to the incidence of particular patterns of group conduct and of individual variations from such patterns. As would be expected, activity in these situations was predominantly concerned with the job and the requirements of the job determined the distribution of time between material and self activities. Consistency of behavior appeared to be a function both of the category and of the individual. Doubling the length of the sample materially increased the self correlation of the data on work-relevant behavior and increased to a slight extent the correlations in all but one of the non-job categories.

To succeed in defining adequately, through the method of direct observation, group patterns of behavior and individual variations expressive of personality, in a highly structuralized work situation, we need considerable further study of the reliability of observers, of factors affecting the sampling of behavior, of the significance of certain personal activities, and of the interrelationships between work and non-work behavior.

[1] A few extreme cases in a small group have a great influence in increasing or decreasing the coefficients of correlation, a distortion which, of course, is not lessened by increasing the length of observation. In a preliminary analysis of Sample I a change in the measure of self consistency in one aspect of behavior from .90 ± .03 to .70 ± .07 was produced by the omission of one worker from a group of 22, while a correlation of .88 ± .03 in another category was unaffected by this change in the group.

CHAPTER V

RELIABILITY OF TECHNIQUES MEASURING SOCIAL-MATERIAL-SELF ACTIVITY PATTERNS

If observational data are to be used for defining individual and group behavior patterns, for indicating deviates from group patterns, for measuring the degree of deviation, or for measuring change over a period of time, we must assume that these ends will be met through statistical analysis. It follows that certain requirements must be fulfilled by the techniques used to produce such data. We consider the following requirements absolutely essential,—(1) that the units of observation shall be similar in form and therefore additive, and (2) that the observational error shall be capable of measurement and the limits of such error kept reasonably narrow.

The chapter on the nursery school technique shows the process by which we met the first condition. We defined our observational categories narrowly in terms of simple overt acts of behavior and timed the duration of each act within the appropriate categories. The use of small time intervals, equal for all categories, meant that quantitative values in terms of the several categories could be assigned to every act, that series of such acts were directly comparable with one another, and that the intervals within a category were properly additive. Such a solution is arbitrary, but in accordance with scientific usage, and it produced workable results. The several techniques used in the study of behavior at other age levels have all used this same time interval system.

The second condition has been less easy to meet. We have given no statistical evidence of reliability in the preceding chapters, although our previous publications and current studies have laid great emphasis on measures of reliability. Our reason for not presenting such evidence is that we are forced to admit its tenuous nature due partly to the slight basis upon which we have had to work and partly to certain inconsistencies in our procedure which have made it impossible to make a direct com-

parison of all our studies. Furthermore, the implications of the results of our study of technique *qua* technique in Part II are so important for this whole problem that part of the discussion will have to be postponed until we have presented these results. We shall, therefore, develop this section cursorily, merely indicating the methods we have used to determine reliability (or, conversely, observational error), the general nature of the results obtained, and some of the reasons for our dissatisfaction with these results.

All our methods have had this common feature. We have always tested reliability by having pairs of observers record the same series of observations simultaneously and have then proceeded to measure the degree of agreement or disagreement of particular pairs of observers.

Our first attempts to measure agreement were made by computing coefficients of correlation between the totals obtained in a given category by two observers, the unit upon which the coefficient of correlation was based being either the child (i. e., all records on a given child were totaled) or the five-minute, or longer, sample. The means for the series were also occasionally computed separately for the two observers and the significance of the difference between the means determined. The assumption with regard to the correlation coefficients was that, if they were considerably less than + 1.00, there was evidence that the technique was not functioning properly and was unreliable. The assumption with regard to the difference between the means for two observers was that, if the difference between the means divided by the standard deviation of this difference were greater than 3, it was highly probably that the technique was unreliable. The value of such measures was, of course, purely negative but they performed a real function in the early stages of our observational studies by helping us to weed out obviously unsuitable techniques. If a technique is not efficient to a degree that will give highly correlated results by two simultaneous observers and that will give means of the behavior which are, within the limits of chance errors, the same, then it must be apparent that the technique is a very poor one indeed! Such measures, however, tell nothing of the degree of observational error which could exist to a marked extent even with a very high correlation

and with no difference at all between the means. Barker's[1] use of these measures is illuminating, since it indicates their value at this early stage and also suggests the dangers involved in their use. She had rather extensive data upon which to base her study of reliability. Four observers, working in six pairwise combinations, had made 366 simultaneous observations on 16 children (averaging about 20 records per child) by the technique illustrated in Chapter I.[2] The five-minute record was taken as the unit of comparison. It will be remembered that her technique was thought to produce data differentiating children on the basis of their social-material-self activity patterns. The data upon which she based her correlations were (1) the physical activity of the child observed, the index of which was computed by measuring the lines representing his movements, (2) the number of social contacts made by the child, the index of which was computed by a simple count of the number of persons recorded as entering into the activity of the child, and (3) the length of time of each discrete activity, the index of which was computed by allocating the time-records to what were thought to be comparable activities recorded by each observer for each child.

Her correlation results were as follows:—the six coefficients for physical activity (i. e., distance covered) ranged from .97 to .98; for number of social contacts, the range was from .47 to .80; for time of discrete activities, from .93 to .98. Disregarding everything but these coefficients, we do not need to go any further to realize that the technique is an extremely inefficient indicator of the degree of social activity of these children since two observers cannot get a high correlation even on so simple an index as the mere *number* of persons with whom a given child on a given record made contacts.

Her use of the significance of the difference between the means was a further check. Two of the six differences were greater than 3 on the measure of physical activity, indicating that the observers might be obtaining consistent results through the operation of a constant bias. It should be noted, however, that the lowest of the correlation coefficients on the number of social contacts (.47) was accompanied by the lowest index of signifi-

[1] Barker, *op. cit.*, Ch. II.

[2] See p. 14, *supra*.

cance of the difference between the means, and the two highest correlation coefficients were accompanied by very high indices of the significance of the difference between the means.

Thus, one measure acted as a check upon the other. The two combined immediately raised grave doubts about the technique, which led, as described, to our further experiments. This study, however, also furnishes an example of the danger of depending upon such measures. The agreement of the six pairs of observers on the time of discrete activities was indicated by correlations ranging from .93 to .98 and the difference between the means divided by the standard deviation of the difference was in every case negligible. Yet, when we look into the method of computing the time for discrete activities, we find the following situation, which Barker frankly admits.

> The times secured by each of two observers were paired. If one observer broke an activity into two parts, whereas the other considered it one, the times paired were the sum of the times obtained for the two activities in the one case, with time for the total activity in the other case. This difference between observers in the division of activities causes some difficulty in determining true reliability. Let us give an example:
>
> *First Observer* (re Edward)
>
> Activity AB—Lies on Beatrice's truck. She pulls him but protests. He pulls truck away from her. Holds it and pats dog................92″
>
> *Second Observer* (re Edward)
>
> Activity AB—Playing with auto truck with Beatrice..................71″
> Activity A′B′—Holds truck and pats dog..............................20″
>
> The reliability results were determined by combining the ninety-two seconds of Observer I with the ninety-one seconds of Observer II. If Observer I had timed the activity with the dog separately the reliability might have been quite different.[1]

It is clear that neither of these measures gives any indication of the fact that the relationship shown may be wholly spurious and that they actually cover up danger-points in the technique.

Recognizing that these measures had only a negative value in determining reliability, we next attempted to derive some simple expression of the amount of interpretive disagreement in a given category and to relate this amount to the total behavior recorded in the category, i. e., to compute a percentage of disagreement.

[1] Barker, *op. cit.*, p. 20.

Some of our difficulties in this respect are illustrated in Loomis' study.

Loomis had recorded physical contacts in fifteen-minute samples. The discrete contacts made by a given child were recorded consecutively in descriptive categories with an indication of the child's rôle in the contact (subject or object) and the identity of the other person involved. This sort of technique led to considerable difficulty in computing any index of disagreement. Three observers had made simultaneous observations in two pairwise combinations, each pair recording for about two hours. The results were analyzed in several ways. The simultaneous observers recorded almost identically the same *number* of contacts during their total period of observation. There was, however, no precise way of determining the identity of contacts recorded by the two observers, since these contacts were recorded discretely and consecutively and were not placed upon any time scale within the fifteen-minute sample, and since, further, one source of observer-disagreement lay in the fact that each of the observers apparently classified as contacts behavior which the other observer either did not see or did not so classify. To measure the extent of classificatory disagreement, contacts recorded by the two observers were assumed to have been the same if the identity of the person contacted was the same on both records and if the contact appeared in approximately the same position within the fifteen-minute sample. On this basis, only 78% of the contacts recorded by one of the two pairs were identifiably the same and 83% of the contacts recorded by the other pair,[1] or, conversely, 22% of the contacts in one case and 17% in the other had been recorded by only one of a given pair of observers. Taking only those contacts definitely identified on the two records as a base, the observers were found to disagree on the rôle of the child observed by 7% and 11%, to disagree on the assignment of contacts to the descriptive categories of initiation (hit, push, pull, caress, etc.) by 22% for each of the pairs, and to disagree on the type of response to the initiated contact

[1] Loomis overestimated the disagreement by selecting as her base the sum of the contacts recorded by the two observers (counted only once) and those recorded by each separately. Obviously the base should have been the sum of all contacts recorded on the two records, i. e., agreements should have been counted twice. The percentages given here are corrected for this error.

by 5% and 9%. It was difficult to decide the true extent of the observational error, but it was obviously large, particularly in the assignment of contacts to descriptive categories and in the decision of the observers as to exactly what represented a discrete contact, since the large proportion of contacts recorded by only one of the two observers in a pair must have been due partly to the fact that a given act was apparently considered as a single contact by one observer and as a succession of contacts by the other observer.

The introduction of the type of technique where contacts were not separated into discrete entities but were recorded as occurring within a given five-second interval made possible a much more precise determination of the extent of intrepretive disagreement. This is illustrated in Arrington's nursery school study.[1] Four observers worked in six pairwise combinations, three pairs making a total of 150 five-minute records on the material-self activities of 20 children (as represented in Form A, page 3, *supra*), the three other pairs recording an equal number on the social-self activities (as represented in Form B, page 3, *supra*). Arrington used both the negative tests (correlations and significance of differences between means) and the more positive test of a percentage of disagreement and agreement.

Each of the following categories was analyzed separately for agreement between pairs of observers: (1) total time spent in contact with material; (2) time spent in inactive contact with material (no overt bodily activity); (3) time spent in physical activity, exclusive of that involved in use of material; (4) time spent in no overt bodily activity (without contact with material); (5) frequency of physical contacts; (6) frequency of talking; (7) frequency of laughing; and (8) frequency of crying. The sub-categories, subject and object physical contacts (those initiated and those received), talking to persons, talking to self, and vocalizing were also analyzed separately.[2]

Her correlations were as follows:—coefficients of .98 for each pairwise combination on *Contact with Material;* .96, .96, .92 for *No Overt Activity* (*in Contact with Material*); .97, .97, .98 for *Physical Activity;* .88, .92, .96 for *No Overt Activity* (*No Contact with Material*); .85, .92 and .95 for *Physical Contacts;* .95, .95, and .96 for *Talking;* .95, .95, .96 for *Laughing;* .93 and .99 for *Crying*

[1] Arrington, *op. cit.*, Ch. IV.

[2] Arrington, *op. cit.*, pp. 30–31.

(one pair of observers recorded crying on only two of fifty records, therefore no coefficient was computed). This negative test, therefore, raised no doubts as to the satisfactory nature of the technique, for all except two of the coefficients of correlation were well over .90 and these were .88 and .85.

With regard to the significance of the difference between the means, one pair of observers had an index greater than 3 in *Talking to Persons*, the same pair also had an index greater than 3 in *Talking to Self*, indicating a possible biased record. All other differences were, in terms of this test, insignificant.

To summarize, the percentages of disagreements could be reasonably determined much more precisely on a time scale than in the other types of studies, but a complicating factor arose in the tendency of certain observers to misplace items by one interval on the time scale as compared with other observers. Therefore, a leeway of one interval was allowed in computing percentages of agreement and adjacent intervals of the same category were considered to be in agreement. The percentage disagreements on this basis were as follows: 1% for *Contact with Material;* 21–22% for *No Overt Activity* (*in Contact with Material*); 9–11% for *Physical Activity;* 21–29% for *No Overt Activity;* 7–12% for *Total Physical Contacts;* 9–12% for *Total Talking;* 17–22% for *Laughing;* 11% for *Crying.*

Some of these percentages of error are very high but most of them cluster around 10% of a given category. We accepted this clustering around 10% as the best possible result we could obtain under the conditions then existing and called the technique "satisfactory" in respect to a category if the proportion of disagreement did not extend much beyond this percentage.

The comparable kindergarten technique showed, on the basis of 100 five-minute records (50 for *Material-Self-Person*, 50 for *Language-Physical Contacts*) made simultaneously by two of our observers, the following range of error:—3% disagreement on *Job Material*, 28% on *Job Self*, 22% *on Non-Job Material*, 8% on *Non-Job Self*, 5% on *Language* and 21% on *Physical Contacts.*

The tests of reliability of the industrial technique were incomplete, being made, as noted on page 81, *supra*, in another situation than the one primarily studied. They showed very small

percentages of errors for the only categories which could be adequately studied at that time: 0.5% in the interpretation of material (*Job* and *Non-Job* combined) and 3–5% in the interpretation of self activity (*Job* and *Non-Job* combined).

We will revert to these tests of reliability after we have discussed our findings in Part II. The following points have seemed to us tenuous and unsatisfactory and have not been solved by our analysis.

(1) The development of a timing technique had obviously produced two sorts of disagreements between observers,—(a) disagreement in assigning a given act of behavior to the same time interval and (b) disagreement in assigning a given act of behavior to the same category. These two kinds of disagreements obviously influenced each other, were not completely separable and complicated our concept of observational error.

(2) We had some evidence of the existence of bias in certain of our observers. This could not be measured and allowed for in our usual comparisons between pairs of observers.

(3) We had no possibility of measuring the varying degrees of self consistency of our several observers or of evaluating this factor in relation to reliability.

(4) We had some evidence that some situations were much more productive of observational errors than others. We had, however, no way of measuring the degree to which this was true and, therefore, our sampling of situations for studying observer-reliability may have been unrepresentative.

PART II

SOME ASPECTS OF OBSERVER RELIABILITY

CHAPTER VI

THE USE OF OBSERVATIONS OF FILM CHARACTERS FOR MEASURING VARIOUS ASPECTS OF OBSERVER RELIABILITY

THE relatively precise determination of activity patterns, described in Part I, was accomplished through a technique of direct observation. The essential features of this technique were the classification of overt acts into predetermined categories and the allocation of these categorically defined acts along a time scale. The degree of precision is dependent to some extent upon the structure of the technique, which was designed to produce results accurate to "within five seconds" for the recording of discrete categories and accurate to "between one and five seconds" for the recording of continuous, mutually exclusive categories. But, however carefully such a technique may have been structuralized *a priori* to produce data of any particular degree of accuracy, its effectiveness will ultimately depend upon the extent to which this structuralization controls the variability among observers recording the same situation and among all observers recording different situations.

If the technique is efficient, the observational records of trained observers in many situations should agree in timing the behavior and in classifying the behavior. The degree of precision or efficiency could be measured in different situations of known degrees of complexity (a) if we could obtain records of agreements and disagreements between simultaneous observations of the same behavior by different observers, or agreements and disagreements between consecutive observations of the same behavior by the same observer, and of agreements and disagreements between empirical observations and records of a known greater degree of precision by the individual observers, (b) if we could disentangle from these records the timing errors on the one hand and the interpretation errors on the other, and (c) if we could, further, allocate the proportion of total observational unreliability due to variation among observers to variation within

each observer and to variation among the behavioral situations observed.

It was clearly impossible to obtain any but the crudest records for measuring the degree of precision of our techniques under ordinary conditions of observation. Agreement between simultaneous observers of the same situation was the only possible means, and it was usually impractical to have more than two observers recording the same situation. No measures of the observer's self consistency were possible, since the observations were of ever-changing situations. The ultimate standard by which the occurrence was checked was the agreement of human observers and there was no record of known greater precision against which their fallibility could be measured.

Moving pictures offered a partial solution to these difficulties. As many observers as desired could observe and record the same situation at the same time. The situation could be reproduced in the same form as often as necessary and consecutive observations could be obtained from the same observers. Observations of the so-called "slow motion" moving picture or even of stilled frames would presumably make possible a record of greater precision against which the crude observations could be checked. With the hope of disentangling the several factors which contributed to the lack of precision in our observations, therefore, we decided to attempt to measure the degree of precision of the technique through observations of the behavior of characters in moving picture films. We made several series of observations, using both ordinary commercial films and a film specially prepared for our purposes. In some of these series, we used a technique comparable to Form B in the Arrington technique[1] in which discrete items of behavior were recorded in several categories. In others, we used a technique for recording behavior in three mutually exclusive categories, and the record was, by definition, continuous.

We describe and analyze the results of two of these series in this volume. That we have not thoroughly solved our problems through this method will be clear. This was partly due to mechanical difficulties, as, for example, the fact that the films were rarely run twice at exactly the same speed and that the situation

[1] See p. 3, *supra.*

was, therefore, not exactly the same at two consecutive observations; the fact that these measurements of observational error demanded the exclusion of extraneous mechanical variation in stop-watches, which was difficult to accomplish; the fact that the perspective of observation could not be made exactly the same for simultaneous observers or reproduced exactly in consecutive observations; the fact that light and shadow varied from one performance to another, that there was some variation in the margin of the films from one performance to another, etc. Our failure to solve all of our problems was also partly due to the unsuitability of existing films for testing techniques applicable to the study of child behavior, and to the prohibitive cost of specially prepared films directly suitable for our purposes. It was, however, to an even greater extent due to the fundamental impossibility of disentangling the several sorts of error and variability with the limited observations we had made. That it has been possible through these observations to throw a great deal more light on the factors involved in the reliability of a technique of this sort than has been possible by any other means, however, makes relevant the presentation of these results.

A description of the two series used in this analysis, the technique, categories, and conditions of observation, follows.

The Moving Picture Series, Rules of Observation and Definitions

The principal body of material for the study of the reliability of the observer and of the technique of observation consisted of records of the behavior of a main character in each of a number of moving picture films. Two series of films (A and C) form the basis of the analysis in the following chapters. The observations of these series were made during 1931 and 1932 by four observers, W, X, Y and Z. A fifth observer, V, made some records for Series C, but not enough to warrant inclusion of the records in the main analysis.[1]

The films comprising Series A, with the numerals used to

[1] Series B, the observations of which were intermediate between those of Series A and C consisted of six films, similar to those in Series A. Mechanical aids to recording were used to control the timing error for certain of the observations. They were also used for half the observations of Series C. These so-called "machine" observations of Series B, will be analyzed, and a complete description of the methods and results will be given in Volume II.

designate them in the following chapters, the length of the observations and the names of the characters observed, were:

Number	Name	Approximate Running Time	Character Observed
I	This Reckless Age	73 minutes	Frances Dee
II	No One Man	71 "	Carole Lombard
III	The Man I Killed	74 "	Phillips Holmes
IV	Tomorrow and Tomorrow	77 "	Ruth Chatterton
V	Shanghai Express	81 "	Marlene Dietrich
VI	Wayward	71 "	Nancy Carroll

Three observations on three consecutive days were made of each of the films.

Series C consisted of records of the behavior of four characters in special films. Each of these films was in duplicate, the pictures having been taken simultaneously by two cameras, one of which was timed to produce four times as many frames per second as the other. The projection of this longer film at the same rate of speed as the shorter gives what is known as a "slow motion" moving picture. This series, therefore, has two parts, which will be distinguished as (1) Series C, normal speed, and (2) Series C, slow speed. Two pictures were taken, four characters appearing in the first and two characters in the second. Of these, three characters (Philip, John and Billy) in the first picture were each observed separately and one (Emma) was observed in the second. The normal speed film of the first picture required approximately 2′ 45″ for projection and the slow speed film of the same picture required four times as long, or 11′. Similarly, the normal speed film of the second picture required approximately 2′ 35″ for projection, but the slow speed film which should have taken 10′ 20″ actually required 10′ 40″ on account of the lack of complete synchronization of the cameras. Consequently, the records of the behavior of Emma could not be used in certain phases of the analysis, such as the comparison of normal and slow frequencies, which is described in Chapter IX.

Because of photographic tricks, the following rules were made to determine when to record. Behavior is recorded

(1) when the character is present and recognizable on the screen,
(2) when he has been clearly recognized in a preceding scene and the view is then limited to a recognizable part of his body (e. g., a

"close-up" of the character's hands and violin shown immediately after a scene where he has been seen playing the violin),

(3) when his image is reflected in a mirror,

(4) when he is seen in a dream or vision or through the projected imagination of another character (but when a "real" scene is superimposed on such a "vision," the behavior in the "real" scene only is recorded), and

(5) when he is seen in a fading scene, until he is completely lost from sight (but when one recordable "real" scene is fading and another such is coming on, the record of the second begins as soon as the character is clearly recognizable in it).

No record of behavior is made

(1) when the character is entirely absent from sight even though his voice is heard, and

(2) at the beginning of the film or of a new scene when the character is unidentifiable because only part of his body is seen or because he is in disguise. In such a case the record in subsequent observations does not begin until the point at which the character was completely recognizable in the first observation, even though the observer has become familiar with the situation and now recognizes the character from memory.

In these observations the overt behavior of characters in moving pictures was classified in terms of three mutually exclusive categories—*Material*, *Self* and *Person*. The terminology is that of the techniques used in the kindergarten and in the factory but the definitions for both these situations were modified to fit the moving picture laboratory situation. A preliminary period of observation preceding Series A served to reveal inconsistencies and lack of precision in the definitions and resulted in their successive refinement. The definitions finally applied in the series analyzed here were as follows:

Material Behavior

The behavior of the character is assigned to the *Material* category when he is seen in functional contact with any material object or when such contact can be legitimately inferred from sound or movement. Functional contact with material includes

(1) manipulation, that is, purposeful change in form or position of an object during its use or exploration by the person observed, and

(2) transportation or change in location, as in picking up, putting down, throwing away, or carrying an object.

The decisive point is the person's movement of the object. The least observable functional movement as, for example, opening a purse or searching a pocket, classifies the behavior as *Material.* If the person's movement of an object is seen or heard, even though the point of contact is not seen, the contact can be legitimately inferred and the behavior is classified as *Material.* This is frequently the case in the unlocking, opening and closing of doors.

Self Behavior

The *Self* category of behavior includes all behavior where there is no functional physical contact either with a material or a person. Non-functional contact with material includes

(1) support or any other contact with material unaccompanied by movement of the material,
(2) non-functional movement of objects, as waving a pencil while talking, and
(3) carrying of objects comparable to clothing, as pocketbook, cane, cigarette, etc. (but taking up, putting down, and other manipulative movements of such objects are recorded under *Material.*

Non-functional contact with persons includes

(1) contact with a person which is inferred but not seen, as in an apparent handshake below the margin of the screen, and
(2) casual contact with a person which does not modify the behavior of either (Example 9, *infra*).

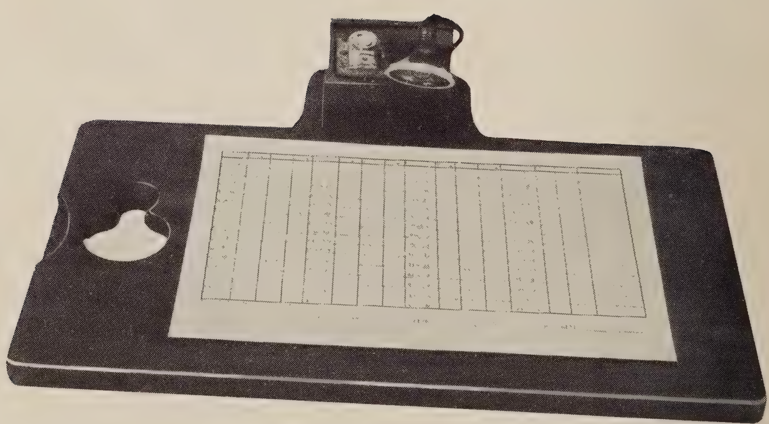
Person Behavior

The *Person* category of behavior includes

(1) direct physical contact with a person or persons when there is no simultaneous functional contact with material (Example 6, *infra*), and
(2) contact with a person or persons through material when the material is not being used functionally (Example 5, *infra*).

Unlike contact with material, contact with a person must be seen to be recorded as such. When an observed personal contact moves off the margin of the screen, the persons being still visible, but the contact no longer in sight, the record of behavior is changed to *Self.* Margins of pictures may differ slightly in suc-

Telechron Clock



Recording Board with Blank Record Form

cessive showings and it is therefore possible for behavior to be correctly recorded as *Person* at one showing and for the same behavior to be correctly recorded as *Self* at another.

The following typical examples of behavior noted in moving pictures are classified in terms of two interacting characters, A and B, according to the above definitions:

Behavior	Category if Observing A	Category if Observing B
1) A lays wrap on B's shoulders	Material	Person
2) A extends object toward B	Material	Self
3) A has object in hand, B reaches for it	Self	Self
4) A takes object, B passive	Material	Person
5) A touches flower worn by B	Person	Person
6) A supports B, who is passive	Person	Person
7) A puts glass to B's lips	Material	Person
8) A holds glass, supporting B who drinks from glass	Person	Material
9) A brushes momentarily against B in crowd, no reaction from B	Self	Self
10) A stumbles into B and knocks him down	Person	Person

Method of Recording

The method of recording used in the series described in this volume was comparable to that used in life situations. Each observer held a small drafting board about 14¼″ by 8¾″ on his lap. A photograph of the board with blank record form is shown here. A flashlight was mounted in the center of the far side of the board. A niche provided in the left side of the board to hold a stop-watch was not used in these observations, since an electric Telechron clock with illuminated dial was substituted for individual stop-watches. (See accompanying photograph). The clock had been selected at the factory for maximum intensity of illumination and, to secure maximum legibility, the minute and hour hands were removed and only a wide second hand, the pointing end of which had been painted black, was substituted for the original second hand. Observers were so seated that each had an unobstructed view of the clock, although the angle of vision necessarily varied somewhat. In Series A the records were made at the local Paramount Theatre. The clock was connected with the electric current by means of an aisle light socket, placed on a shelf formed by the existing partition between the

second and third rows of seats (the observers sat in the third and fourth rows of a loge) and set at exactly 12:00. One observer kept his hand on the switch to start the clock at the proper moment. Experience in watching the preliminary captions had shown that the title of the picture was always flashed after certain preliminaries. At the first clear flash of the entire title, the observer turned the switch which illuminated the clock dial and set the second hand in motion, and the observation began.

Each observer held on his board consecutively numbered record forms, each of which was blocked off to represent 60 five-second intervals.[1] Four spaces were provided horizontally for each interval. The first column contained the number of the interval, the number of the minute and the number of the second with which the interval began. The remaining three columns were labeled *Material*, *Self* and *Person*. Before the character chosen for observation appeared, the observer followed a routine of looking from the screen to the clock to the paper periodically every five seconds, and checking the time in the correct interval on the paper. As soon as the character appeared and was recognized the observer drew a vertical pencil line, in the appropriate column, through as many intervals as the behavior lasted. Whenever a change in category of behavior occurred, the line was transferred to the appropriate new column. A short horizontal line was drawn at the proper second to indicate an exit of the character or a change of scene, but not a change of perspective in the same scene. Any change of behavior necessitated a glance at the clock and form. In addition to those glances the observer attempted to follow the routine of looking from the screen to the clock to the paper every five seconds. It was agreed that no attempt should be made to record behavior, the duration of which was less than a second. The recording of behavior during the character's presence, and the checking of the time in the intervals when the character was absent, continued until the words "The End" were flashed upon the screen.

The observations of Series C, normal and slow speed, were made in a dark room at the Institute of Human Relations, since the films were of the 16 mm. slow burning type and were suitable for the ordinary home projector. There were several difficulties

[1] See Appendix A, Chart A.

to be overcome before recording could be carried through satisfactorily. The use of direct current produced startling variations in consecutive runnings of the same film. The use of alternating current for both clock and projector improved matters only slightly. Large variations still appeared between morning and afternoon projections and the first projection of a series always took longer than succeeding projections. Even the use of a projector with a synchronous motor failed to give satisfactory results unless the motor was thoroughly warmed by running the projector for abour ten minutes before any film was projected.[1] There were still some slight variations in successive running times but these were probably due in part to inaccurate starting or stopping of the clock. Since there was no title or other preliminary caption, the first horizontal frame containing a picture was selected for the starting point and the last such frame for the stopping point.

After a practice period for all, the observers worked in pairs, one using the usual method throughout an observation period while the other used a mechanical aid to recording, designed to control timing errors. The schedules were so arranged that the "machine" and ordinary methods were alternated for each observation by each observer. Each observation consisted of the normal and slow speed films for two characters. Six normal and six slow records of each of the four characters were made by Observers W, X, Y and Z. Observer V made the six normal and six slow records on only two of the characters. Each observer made three of the normal records and three of the slow records on each character by the ordinary method and the remainder by the "machine" method. The ordinary records are the only ones considered in this volume.

Transcription of Records

In order to permit direct comparison and analysis of simultaneous records these were transcribed to master sheets[2] where the corresponding records of the several observers were plotted in

[1] These details are given as an illustration of the compounding of difficulties which the student of behavior finds when he begins to depend upon machines as an aid to observation. Unless the error in the machines can be reduced to very narrow limits, they are worse than useless as an aid to determining the degree of observational error.

[2] The rules for transcription with a detailed example are given in Appendix A.

parallel columns. A master sheet was made for each observation of each film in Series A. To facilitate rapid and accurate transcription a form as nearly like the original record sheet as possible was devised. One master sheet form covered five minutes of the record just as the record form did and there were as many pages of the master sheet as there were of the original record. Five spaces were provided horizontally for each five-second interval, the first containing the chronological number of the interval, the other four to contain the transcribed records of the four simultaneous observers. The spaces representing whole intervals were the same size as those on the record form but on the master sheet they were bisected by horizontal lines into half intervals. A master sheet form could then be cut into columns of half intervals, and these columns used to measure the duration of behavior on the original records.

For Series C, normal and slow speed, one master sheet was made for each series of six consecutive observations of one character by one observer. There were four normal and four slow master sheets for each of the observers W, X, Y and Z. On this master sheet form there were seven horizontal spaces for each of 120 quarter intervals. It was necessary to transcribe these records in quarter intervals to facilitate comparison of normal records with slow, which contained exactly four times as much. The first column was used to number the whole intervals. The remaining columns were used for the first, second and third observations made by the ordinary, and by the machine, methods.

Master sheets comparable to those of Series A were made for the third observation of each character in Series C, normal and slow, where the unit was the half interval and where the records of the several observers were plotted in parallel columns.

Great care was taken in the making of all master sheets. The transcription was done by one person and checked by a second. If they disagreed in a measurement or an interpretation, the point was referred to the recorder for final judgment. Thus accuracy of transcription was well established before any tabulations were made.

CHAPTER VII

THE MEASUREMENT OF TIMING TENDENCIES AND TIMING DISAGREEMENTS

In the process of observing, classifying and recording behavior on a time scale the observer should, ideally, accomplish a fourfold sequence of acts with instantaneous precision. He first identifies the individual to be observed, then classifies the activity of this individual as material, self or person behavior, in accordance with predetermined definitions, then notes on the stop-watch or clock the time at which the activity began, and finally, makes his record in the appropriate column and time interval on the mimeographed record blank. This sequence—identification, interpretation, timing and recording—is repeated in the moving picture observations with each shift of scene or change of behavior within a scene. While each of these acts is distinct in the initial process of recording, the resultant effects of each upon the record can be differentiated only by approximate measures.

Our analysis of timing discrepancies is made, of necessity, on a relative basis, i. e., in terms of the performance of each observer with each other observer as the standard. In this series of observations, it is impossible to know to what extent a given observer is more accurate than another in his timing of the behavior observed. We can only determine the fact of his disagreement with other observers and the direction and amount of this disagreement, and draw certain inferences from such comparisons.

The factors involved in this disagreement must as far as possible be isolated. Three main causes of timing disagreement can be distinguished,—(1) inaccuracy of timing instruments, by which is meant lack of synchronization of stop-watches used by different observers, (2) inherent characteristics of the behavior observed and of the recording method, and (3) inaccuracy of the observer himself in seeing, defining or recording behavior.

The first of these factors does not constitute a serious problem in real life studies of behavior based on short time samples and

has been eliminated in the two motion picture series now under consideration. The importance that this factor could assume in motion picture records became evident in some of our preliminary studies and the findings with regard to this point are presented in Appendix B. This cause of discrepant timing was almost completely eliminated in the series of observations here under discussion by the use of a single time piece for all observers. The electric clock with illuminated dial[1] which was placed in a position central for all observers, automatically ruled out timing differences due to non-synchronous timing instruments.[2]

Presumably the varying degrees of complexity and rapidity of occurrence of behavior in different categories are reflected in timing tendencies and errors. This factor was controlled by analyzing the timing factor separately for each category.

The third factor, the personal element in timing, is reflected in the chance inaccuracies and the persistent biases of the recorder and it is this factor which we attempt to measure in its total effect, although we are unable to weight its components. Individual peculiarities may affect any one of the several processes involved in the taking of a record. The individual's speed of reaction to stimuli may affect the initial act of seeing the events, causing him to accelerate or retard the entire sequence of processes. Varying familiarity with, or consciousness of, definitions affects the speed with which behavior is classified and subsequently recorded. Individuals may also differ in their manner of making the record. One observer, uncertain as to the exact moment at which the behavior began, may consistently place the beginning of his line of record at the beginning of the five-second interval closest to the time noted on the clock, while another, under similar circumstances, may habitually locate his record at the middle or at the end of the interval. Thus the particular way in which a person sees, interprets and records events determines the degree of accuracy with which he times their occurrence. It is with this third source of timing error, the personal idiosyncrasies of the observer, that we are concerned in the discussion of method and results which follows.

[1] See p. 121. *supra.*

[2] There was some evidence that the location of the observer with reference to his vision of the clock had a slight distorting effect on the records, but the data are insufficient to prove the point.

A twofold analysis of the timing factor in reliability has been made,—(1) in terms of the differential timing tendencies of the several observers, and (2) in terms of the proportion of discrepantly timed units of the several observers.

Measurement of Timing Tendencies

Timing tendency, as we define it, is synonymous with skewness when the norm is synchronous recording. We assume, in other words, that a given pair of observers will record most of the items of behavior in corresponding intervals, but that each of them will misplace some items by recording either in the interval preceding or in the interval following that in which the other observer records. We further assume that, in the absence of persistent idiosyncrasies, those items preceding the standard observer's record will tend to balance those following the standard observer's record, and that the timing tendency of each observer, as compared with other observers, can be indicated by his persistent misplacement of items in a particular direction.

The first step in the timing analysis was the transcription of the records of the four observers for each observation from the original record blanks to a single master sheet.[1] Tabulations by film, observation and category were then made of all the time units[2] in which any observer had a record of behavior. Each unit was tabulated in terms of the particular combination of observers recording in that unit, that is, synchronously, the combination recording the behavior in the preceding unit, and the combination recording in the following time unit. The horizontal spaces on these tabulation forms[3] represent the records which we have called synchronous. Two strokes were made in these spaces for each time unit, one in the + column corresponding to the exact combination of observers recording in the preceding unit, and one in the − column corresponding to the combination recording in the following unit. The symbols + and − at the top of the vertical columns indicate respectively "preceded by" and "followed by" and were later translated in terms of retardation and acceleration. The totals of the hori-

[1] See pp. 123–124, *supra*, and Appendix A.
[2] The term "unit" will be used for the 2½″ time interval, "interval" for the 5″ interval.
[3] See Appendix C, Form A.

zontal rows were divided by 2 to obtain the actual number of units in which records were made. The vertical columns were totaled separately for +'s and −'s. From this first comprehensive tabulation form separate tabulations were made for each pair of observers. This step consisted in the transference of totals for each combination in which one or both observers in a given pair appeared on Tabulation Form A to the appropriate spaces on Tabulation Form B[1] for the given pair.

The method of weighting used in computing indices of timing tendency is based on the assumption that the larger the number of observers recording an event in a given time unit, the greater is the probability that the event occurred in that unit. It also assumes that any record, even though apparently synchronous with the records of other observers, may actually be incorrectly timed, that is, it may have become synchronous by virtue of its acceleration or retardation. Each time unit in which an observer made a record of a given type of behavior is therefore considered in the light of its probable accurate timing, its probable acceleration and its probable retardation, as determined by the number of persons recording in the same unit, in the unit following, and in the unit preceding, respectively. In pairwise comparison of any two observers, a single record of one observer preceded by no record and followed by a record by both observers composing the pair receives a weight of 1 for the "synchronous" record, 2 for acceleration and 0 for retardation. A single record preceded by a record by both members of the pair and followed by no record is weighted 1 for the "synchronous" record, 2 for retardation and 0 for acceleration. When the two observers record synchronously, each receives a weighting of 2.

All recorded units were tabulated and weighted separately for each pair of observers and each category. The weighted "synchronous," retarded and accelerated units were totaled separately. The tendency toward acceleration or retardation for a given observer in a pair was expressed as a percentage by dividing the difference between the observer's weighted + and weighted − units by the average number of weighted "synchronous" units for the pair. The tendency toward acceleration or retardation of a given observer compared with all other

[1] See Appendix C, Form B.

observers was expressed as a percentage by dividing the difference between his total number of weighted + units for all pairs combined and his total number of weighted − units for all pairs combined by the average number of weighted "synchronous" units per pair for all pairs combined.[1] These percentages, it is clear, do not constitute a measure of total amount of timing discrepancy, but are indices of differential tendencies on the part of observers, when compared with other observers, to have more retarded units than accelerated, or more accelerated units than retarded. If an observer has no timing tendency, it is to be expected that the number of units which, by chance, he retards, will balance the number he accelerates. It is the excess of accelerated over retarded or of retarded over accelerated units which we have attempted to measure in the timing tendencies discussed in this chapter.

The method just described was used in all computations of timing indices for individual observers. Since each unit was counted twice, once in terms of acceleration and once in terms of retardation, the total amount of retardation for a group of observers (two, three, four or more persons) balances the total amount of acceleration for the group. Consequently, the sum of the percentages of the retarded observers equals the sum of the percentages of those who are accelerated. This balance ± figure may be used, therefore, as an index of the amount of timing discrepancy for the group.

The percentage indices of timing tendency for Observers W, X, Y and Z per film, observation and category for Series A, based on pairwise comparison of each observer with each other observer, are presented in Table XI. Percentages per observation, based on all first, all second and all third observations of the six films combined, and percentages per film, based on the three observations of each film combined, for each observer in each pairing, are included in this table. These figures represent the average percentage of timing discrepancy, not the average of the distribution of percentages of timing discrepancy. The corresponding figures per film, observation and category for the combined data for three pairings of each observer with each other

[1] See Appendix C.

TABLE XI

Percentage Indices of Timing Tendency per Observer Expressing the Ratio Between the Number of Retarded Units Exceeding the Accelerated, or the Number of Accelerated Units Exceeding the Retarded, and the Average Number of "Synchronous" Units Per Pair, by Film, Observation, Category of Behavior and Pairwise Combination of Observers (Series A) *

— = Acceleration + = Retardation 0 = No excess in either direction

Standard Observer	Co-observers	Observation	Material							Self							Person						
			Films							Films							Films						
			I	II	III	IV	V	VI	I–VI	I	II	III	IV	V	VI	I–VI	I	II	III	IV	V	VI	I–VI
W	X	1	−2.7	+4.9	−9.5	−4.4	−1.5	−1.2	−1.5	+0.6	−0.9	−0.9	−1.0	−1.9	−1.3	−1.0	−0.9	−0.2	+2.8	−1.5	−1.3	−2.0	−0.8
		2	+1.4	−0.5	−1.4	−6.4	−5.6	+2.6	−1.2	−1.2	−0.9	+0.8	−0.5	−1.5	−0.1	−0.6	0	−1.0	+1.6	+0.7	−5.0	−5.1	−1.4
		3	−4.3	+0.5	−5.5	−3.4	−2.4	+3.6	−0.7	−2.4	+0.1	−0.3	−0.7	−1.6	−0.1	−0.7	−11.3	+1.2	+1.7	−0.6	−7.5	+0.5	−1.6
		Av.	−1.8	+1.4	−5.3	−4.7	−3.1	+2.0	−1.1	−1.0	−0.6	−0.1	−0.7	−1.6	−0.5	−0.8	−4.2	+0.1	+2.0	−0.5	−4.7	−2.2	−1.2
	Y	1	+0.7	0	−11.8	+0.6	−1.0	−5.7	−2.0	+0.9	−0.2	−0.9	−0.7	−1.0	−1.4	−0.6	−0.8	−2.4	+2.3	−1.4	−1.7	−6.8	−2.3
		2	+3.6	−3.0	+7.2	−7.1	−4.0	+0.6	−1.2	+1.2	−1.1	+0.8	−0.6	−0.4	+0.2	−0.1	+1.1	−1.3	+1.1	−1.5	−3.3	−0.5	−0.8
		3	−8.3	−2.6	+3.7	−2.1	−1.7	+0.3	−1.9	−1.4	−0.9	−0.3	−0.4	−0.6	+0.5	−0.4	−6.3	+0.8	0	+1.3	−0.6	+1.0	−0.2
		Av.	−1.5	−2.0	+0.4	−3.0	−2.2	−1.0	−1.7	+0.2	−0.8	−0.1	−0.6	−0.7	−0.2	−0.4	−2.1	−0.9	+1.1	−0.6	−1.9	−2.0	−1.1
	Z	1	+19.5	+8.2	+12.1	+7.3	+6.0	+4.3	+8.2	+4.4	+2.7	+2.2	+2.6	+3.3	+2.4	+2.8	+5.3	+6.4	+2.3	+4.9	+6.5	+1.3	+4.6
		2	+12.2	+5.1	+13.9	+10.2	+2.2	+4.7	+6.5	+4.4	+1.8	+2.9	+2.2	+2.8	+2.6	+2.7	+10.6	+4.1	+3.1	+2.6	+5.4	+2.3	+4.3
		3	+7.4	+2.3	+8.0	+1.0	−3.3	+5.7	+2.9	+3.3	+2.4	+1.8	+1.3	+3.2	+2.0	+2.2	+2.6	+6.1	+1.7	+4.2	+3.6	+3.7	+4.2
		Av.	+12.6	+4.9	+11.3	+5.9	+1.5	+5.0	+5.7	+4.0	+2.3	+2.3	+2.0	+3.1	+2.3	+2.6	+6.2	+5.5	+2.4	+3.9	+5.2	+2.5	+4.3
X	W	1	+2.7	−4.9	+9.5	+4.4	+1.5	+1.2	+1.5	−0.6	+0.9	+0.9	+1.0	+1.9	+1.3	+1.0	+0.9	+0.2	−2.8	+1.5	+1.3	+2.0	+0.8
		2	−1.4	+0.5	+1.4	+6.4	+5.6	−2.6	+1.2	+1.2	+0.9	−0.8	+0.5	+1.5	+0.1	+0.6	0	+1.0	−1.6	−0.7	+5.0	+5.1	+1.4
		3	+4.3	−0.5	+5.5	+3.4	+2.4	−3.6	+0.7	+2.4	−0.1	+0.3	+0.7	+1.6	+0.1	+0.7	+11.3	−1.2	−1.7	+0.6	+7.5	−0.5	+1.6
		Av.	+1.8	−1.4	+5.3	+4.7	+3.1	−2.0	+1.1	+1.0	+0.6	+0.1	+0.7	+1.6	+0.5	+0.8	+4.2	−0.1	−2.0	+0.5	+4.7	+2.2	+1.2
	Y	1	0	−5.2	+5.6	+3.4	0	−3.2	−0.5	+0.3	−0.2	+0.1	+0.2	+1.3	−0.5	+0.1	−0.4	−1.5	−1.1	+1.8	−1.0	−5.5	−1.7
		2	+1.8	+2.0	+7.9	−2.9	−2.6	+2.2	+0.9	+1.6	−0.2	+0.2	+0.4	+1.2	+0.8	+0.6	+2.3	0	−1.0	−0.2	+1.1	+2.4	+0.7
		2	+1.9	−3.9	+9.6	+2.2	0	0	+0.6	+1.3	−0.6	−0.3	0	+0.7	+0.6	+0.2	+1.9	−0.5	−0.6	+1.4	+5.5	+0.5	+0.9
		Av.	+1.3	−2.1	+7.8	+0.9	−0.8	0	+0.4	+1.1	−0.4	−0.0	+0.2	+1.1	+0.3	+0.3	+1.3	−0.7	−0.9	+0.1	+2.9	−0.7	0
	Z	1	+13.1	+4.9	+20.9	+12.7	+8.1	+2.9	+8.5	+3.5	+3.6	+2.5	+3.4	+4.5	+2.8	+3.4	+8.4	+7.1	−1.1	+7.1	+6.8	+3.7	+5.8
		2	+11.5	+6.9	+8.8	+13.5	+6.3	+2.9	+7.3	+4.9	+2.7	+2.5	+2.9	+4.9	+2.5	+3.2	+12.5	+5.2	+0.5	+2.8	+5.6	+6.1	+5.3
		3	+11.5	+1.5	+9.2	+6.5	−0.8	+4.4	+4.3	+3.9	+2.3	+1.9	+1.8	+4.7	+2.3	+2.7	+8.0	+4.3	+2.3	+3.8	+6.4	+4.6	+4.7
		Av.	+12.0	+4.3	+12.8	+10.7	+4.2	+3.5	+6.5	+4.1	+2.9	+2.3	+2.7	+4.7	+2.5	+3.1	+9.7	+5.5	+0.5	+4.6	+6.3	+4.8	+5.3

* All units weighted by number of persons recording.

TABLE XI—*Continued*

Percentage Indices of Timing Tendency per Observer Expressing the Ratio Between the Number of Retarded Units Exceeding the Accelerated, or the Number of Accelerated Units Exceeding the Retarded, and the Average Number of "Synchronous" Units Per Pair, by Film, Observation, Category of Behavior and Pairwise Combination of Observers (Series A) *

— = Acceleration   + = Retardation   0 = No excess in either direction

Standard Observer	Co-observers	Observation	Material							Self							Person						
			Films							Films							Films						
			I	II	III	IV	V	VI	I–VI	I	II	III	IV	V	VI	I–VI	I	II	III	IV	V	VI	I–VI
Y	W	1	−0.7	0	+11.8	−0.6	+1.0	+5.7	+2.0	−0.9	+0.2	+0.9	+0.7	+1.0	+1.4	+0.6	+0.8	+2.4	−2.3	+1.4	+1.7	+6.8	+2.3
		2	−3.6	+3.0	−7.2	+7.1	+4.0	−0.6	+1.2	−1.2	+1.1	−0.8	+0.6	+0.4	−0.2	+0.1	−1.1	+1.3	−1.1	+1.5	+3.3	+0.5	+0.8
		3	+8.3	+2.6	−3.7	+2.1	+1.7	−0.3	+1.9	+1.4	+0.9	+0.3	+0.4	+0.6	−0.5	+0.4	+6.3	−0.8	0	−1.3	+0.6	−1.0	+0.2
		Av.	+1.5	+2.0	−0.4	+3.0	+2.2	+1.0	+1.7	−0.2	+0.8	+0.1	+0.6	+0.7	+0.2	+0.4	+2.1	+0.9	−1.1	+0.6	+1.9	+2.0	+1.1
	X	1	0	+5.2	−5.6	−3.4	0	+3.2	+0.5	−0.3	+0.2	−0.1	−0.2	−1.3	+0.5	−0.1	+0.4	+1.5	+1.1	−1.8	+1.0	+5.5	+1.7
		2	−1.8	−2.0	−7.9	+2.9	+2.6	−2.2	−0.9	−1.6	+0.2	−0.2	−0.4	−1.2	−0.8	−0.6	−2.3	0	+1.0	+0.2	−1.1	−2.4	−0.7
		3	−1.9	+3.9	−9.6	−2.2	0	0	−0.6	−1.3	+0.6	+0.3	0	−0.7	−0.6	−0.2	−1.9	+0.5	+0.6	−1.4	−5.5	−0.5	−0.9
		Av.	−1.3	+2.1	−7.8	−0.9	+0.8	0	−0.4	−1.1	+0.4	+0.0	−0.2	−1.1	−0.3	−0.3	−1.3	+0.7	+0.9	−0.1	−2.9	+0.7	0
	Z	1	+9.4	+6.0	+10.8	+7.3	+6.9	+9.2	+8.0	+4.3	+3.7	+2.7	+3.1	+3.7	+3.2	+3.4	+6.1	+6.9	0	+7.4	+5.7	+6.2	+6.1
		2	+13.3	+9.0	+9.1	+15.5	+10.1	+5.5	+9.8	+4.1	+3.0	+2.6	+2.2	+3.2	+2.2	+2.8	+7.0	+4.3	+2.0	+3.1	+3.7	+4.1	+4.1
		3	+13.9	+5.3	+1.2	+5.0	+0.4	+5.6	+5.3	+3.4	+2.8	+2.2	+2.0	+3.2	+1.5	+2.5	+6.5	+4.4	+2.7	+4.0	+3.9	+4.2	+4.4
		Av.	+12.2	+6.8	+6.8	+9.1	+5.5	+6.5	+7.6	+3.9	+3.2	+2.5	+2.4	+3.4	+2.3	+2.8	+6.6	+5.2	+1.6	+4.8	+4.4	+4.8	+4.8
Z	W	1	−19.5	−8.2	−12.1	−7.3	−6.0	−4.3	−8.2	−4.4	−2.7	−2.2	−2.6	−3.3	−2.4	−2.8	−5.3	−6.4	−2.3	−4.9	−6.5	−1.3	−4.6
		2	−12.2	−5.1	−13.9	−10.2	−2.2	−4.7	−6.5	−4.4	−1.8	−2.9	−2.2	−2.8	−2.6	−2.7	−10.6	−4.1	−3.1	−2.6	−5.4	−2.3	−4.3
		3	−7.4	−2.3	−8.0	−1.0	+3.3	−5.7	−2.9	−3.3	−2.4	−1.8	−1.3	−3.2	−2.0	−2.2	−2.6	−6.1	−1.7	−4.2	−3.6	−3.7	−4, 2
		Av.	−12.6	−4.9	−11.3	−5.9	−1.5	−5.0	−5.7	−4.0	−2.3	−2.3	−2.0	−3.1	−2.3	−2.6	−6.2	−5.5	−2.4	−3.9	−5.2	−2.5	−4.3
	X	1	−13.1	−4.9	−20.9	−12.7	−8.1	−2.9	−8.5	−3.5	−3.6	−2.5	−3.4	−4.5	−2.8	−3.4	−8.4	−7.1	+1.1	−7.1	−6.8	−3.7	−5.8
		2	−11.5	−6.9	−8.8	−13.5	−6.3	−2.9	−7.3	−4.9	−2.7	−2.5	−2.9	−4.9	−2.5	−3.2	−12.5	−5.2	−0.5	−2.8	−5.6	−6.1	−5.3
		3	−11.5	−1.5	−9.2	−6.5	+0.8	−4.4	−4.3	−3.9	−2.3	−1.9	−1.8	−4.7	−2.3	−2.7	−8.0	−4.3	−2.3	−3.8	−6.4	−4.6	−4.7
		Av.	−12.0	−4.3	−12.8	−10.7	−4.2	−3.5	−6.5	−4.1	−2.9	−2.3	−2.7	−4.7	−2.5	−3.1	−9.7	−5.5	−0.5	−4.6	−6.3	−4.8	−5.3
	Y	1	−9.4	−6.0	−10.8	−7.3	−6.9	−9.2	−8.0	−4.3	−3.7	−2.7	−3.1	−3.7	−3.2	−3.4	−6.1	−6.9	0	−7.4	−5.7	−6.2	−6.1
		2	−13.3	−9.0	−9.1	−15.5	−10.1	−5.5	−9.8	−4.1	−3.0	−2.6	−2.2	−3.2	−2.2	−2.8	−7.0	−4.3	−2.0	−3.1	−3.7	−4.1	−4.1
		3	−13.9	−5.3	−1.2	−5.0	−0.4	−5.6	−5.3	−3.4	−2.8	−2.2	−2.0	−3.2	−1.5	−2.5	−6.5	−4.4	−2.7	−4.0	−3.9	−4.2	−4.4
		Av.	−12.2	−6.8	−6.8	−9.1	−5.5	−6.5	−7.6	−3.9	−3.2	−2.5	−2.4	−3.4	−2.3	−2.8	−6.6	−5.2	−1.6	−4.8	−4.4	−4.8	−4.8

* All units weighted by number of persons recording.

TABLE XII

*Percentage Indices of Timing Tendency per Observer Expressing the Ratio Between the Sum of All Retarded Units Exceeding the Accelerated, or of All Accelerated Units Exceeding the Retarded, in Three Pairings with Other Observers, and the Average Number of "Synchronous" Units Per Pair in Three Pairings with Other Observers, by Film, Observation, and Category of Behavior, with Corresponding Group Indices Representing the Sums of the Percentages of the Four Observers (Series A)**

— = Acceleration + = Retardation 0 = No excess in either direction

Standard Observer	Co-observers	Observation	Material							Self							Person						
			Films							Films							Films						
			I	II	III	IV	V	VI	I–VI	I	II	III	IV	V	VI	I–VI	I	II	III	IV	V	VI	I–VI
W	XYZ	1	+5.3	+4.4	−3.0	+1.2	+1.3	−0.6	+1.6	+2.0	+0.5	+0.2	+0.3	+0.1	−0.2	+0.4	+1.2	+1.2	+2.4	+0.6	+1.2	−2.5	+0.5
		2	+5.5	+0.5	+6.7	−1.3	−2.4	+2.7	+1.4	+1.4	−0.1	+1.5	+0.4	+0.3	+0.9	+0.6	+4.1	+0.7	+1.9	+0.6	−1.0	−1.0	+0.7
		3	−1.9	0	+2.1	−1.4	−2.4	+3.3	+0.1	−0.2	+0.5	+0.4	+0.1	+0.3	+0.8	+0.3	−4.8	+2.7	+1.1	+1.5	−1.5	+1.7	+0.8
		Av.	+2.8	+1.4	+2.1	−0.6	−1.3	+2.1	+1.0	+1.1	+0.3	+0.7	+0.2	+0.2	+0.5	+0.4	+0.1	+1.6	+1.8	+0.9	−0.4	−0.5	+0.7
X	WYZ	1	+4.7	−1.8	+11.8	+6.8	+3.3	+0.6	+3.2	+1.1	+1.4	+1.2	+1.5	+2.5	+1.2	+1.5	+2.9	+1.9	−1.6	+2.4	+3.2	+0.2	+1.6
		2	+3.7	+2.9	+5.8	+5.4	+3.3	+0.8	+3.0	+2.6	+1.1	+0.6	+1.2	+2.5	+1.1	+1.5	+4.7	+2.1	−0.7	+0.6	+3.7	+4.5	+2.4
		3	+5.4	−0.9	+8.0	+3.9	+0.5	+0.2	+1.8	+2.6	+0.6	+0.6	+0.8	+2.3	+1.0	+1.2	+6.5	+0.8	0	+2.0	+6.5	+1.6	+2.4
		Av.	+4.6	+0.2	+8.4	+5.4	+2.3	+0.5	+2.6	+2.1	+1.0	+0.8	+1.2	+2.4	+1.1	+1.4	+4.8	+1.5	−0.8	+1.7	+4.5	+2.1	+2.1
Y	WXZ	1	+3.5	+3.6	+5.9	+1.3	+2.6	+5.4	+3.6	+0.9	+1.3	+1.1	+1.2	+1.1	+1.7	+1.2	+2.7	+3.8	−0.4	+3.3	+2.0	+6.1	+3.5
		2	+2.8	+3.6	−2.2	+8.7	+5.4	+0.9	+3.5	+0.4	+1.4	+0.5	+0.8	+0.7	+0.4	+0.7	+1.3	+1.9	+0.7	+1.6	+2.1	+0.8	+1.5
		3	+7.4	+4.0	−4.2	+1.8	+0.7	+1.7	+2.3	+1.1	+1.4	+0.9	+0.8	+1.0	+0.1	+0.9	+4.0	+1.4	+1.1	+0.4	−0.4	+1.0	+1.2
		Av.	+4.6	+3.8	−0.4	+3.9	+2.8	+2.4	+3.1	+0.8	+1.4	+0.9	+0.9	+1.0	+0.7	+1.0	+2.7	+2.4	+0.5	+1.8	+1.2	+2.5	+2.1
Z	WXY	1	−13.5	−6.2	−14.7	−9.3	−7.2	−5.4	−8.4	−4.0	−3.2	−2.5	−3.0	−3.7	−2.7	−3.1	−6.8	−6.9	−0.4	−6.3	−6.4	−3.8	−5.6
		2	−12.0	−7.0	−10.3	−12.8	−6.3	−4.4	−7.9	−4.4	−2.4	−2.6	−2.4	−3.5	−2.4	−2.8	−10.1	−4.7	−1.9	−2.8	−4.8	−4.3	−4.6
		3	−10.9	−3.1	−5.9	−4.3	+1.2	−5.2	−4.2	−3.5	−2.5	−1.9	−1.7	−3.6	−1.9	−2.4	−5.7	−4.9	−2.2	−3.9	−4.6	−4.3	−4.4
		Av.	−12.0	−5.4	−10.1	−8.7	−3.8	−5.0	−6.7	−4.0	−2.7	−2.4	−2.3	−3.6	−2.3	−2.8	−7.6	−5.5	−1.5	−4.4	−5.3	−4.1	−4.9
Group		1	±13.5	±8.0	±17.7	±9.3	±7.2	±6.0	±8.4	±4.0	±3.2	±2.5	±3.0	±3.7	±2.9	±3.1	±6.8	±6.9	±2.4	±6.3	±6.4	±6.3	±5.6
		2	±12.0	±7.0	±12.5	±14.1	±8.7	±4.4	±7.9	±4.4	±2.5	±2.6	±2.4	±3.5	±2.4	±2.8	±10.1	±4.7	±2.6	±2.8	±5.8	±5.3	±4.6
		3	±12.8	±4.0	±10.1	±5.7	±2.4	±5.2	±4.2	±3.7	±2.5	±1.9	±1.7	±3.6	±1.9	±2.4	±10.5	±4.9	±2.2	±3.9	±6.5	±4.3	±4.4
		Av.	±12.0	±5.4	±10.5	±9.3	±5.1	±5.0	±6.7	±4.0	±2.7	±2.4	±2.3	±3.6	±2.3	±2.8	±7.6	±5.5	±2.3	±4.4	±5.7	±4.6	±4.9

* All units weighted by number of persons recording.

observer, together with the balance figures for the group, are shown in Table XII.

Table XIII shows percentage indices of timing tendency based on the total record of behavior, without regard to interpretation, for each observer compared with each other observer separately and with all three observers combined, in the third observation of the six films of series A.

TABLE XIII

*Percentage Indices of Timing Tendency per Observer Expressing the Ratio between the Number of Retarded Units Exceeding the Accelerated, or the Number of Accelerated Units Exceeding the Retarded, and the Average Number of "Synchronous" Units per pair, for the Total Record of Behavior without Distinction of Categories, in Films I-VI, Observation 3, Series A.**

− = Acceleration   + = Retardation   0 = No excess in either direction

Standard Observer	Co-Observer	Films						
		I	II	III	IV	V	VI	I-VI
W	X	−1.8	−0.6	−0.4	−0.8	−1.0	−0.6	−0.8
	Y	−1.1	−0.7	−0.3	−0.6	−0.2	+0.4	−0.4
	Z	+1.7	+1.2	+1.5	+0.9	+2.4	+1.6	+1.5
	XYZ	−0.4	0	+0.3	−0.1	+0.4	+0.5	+0.1
X	W	+1.8	+0.6	+0.4	+0.8	+1.0	+0.6	+0.8
	Y	+0.8	−0.3	+0.1	+0.2	+0.5	+0.9	+0.3
	Z	+2.8	+1.6	+2.0	+1.6	+3.2	+2.0	+2.1
	WYZ	+1.8	+0.7	+0.8	+0.8	+1.5	+1.1	+1.1
Y	W	+1.1	+0.7	+0.3	+0.6	+0.2	−0.4	+0.4
	X	−0.8	+0.3	−0.1	−0.2	−0.5	−0.9	−0.3
	Z	+2.0	+1.8	+1.7	+1.6	+2.7	+1.3	+1.8
	WXZ	+0.8	+0.9	+0.6	+0.7	+0.8	+0.0	+0.6
Z	W	−1.7	−1.2	−1.5	−0.9	−2.4	−1.6	−1.5
	X	−2.8	−1.6	−2.0	−1.6	−3.2	−2.0	−2.1
	Y	−2.0	−1.8	−1.7	−1.6	−2.7	−1.3	−1.8
	WXY	−2.2	−1.6	−1.7	−1.4	−2.7	−1.6	−1.8
Group		±2.6	±1.6	±1.7	±1.5	±2.7	±1.6	±1.8

* All units weighted by number of persons recording.

In discussing the data presented in these tables, the following points will be considered.

(1) What is the extent of timing tendency per category of behavior? Does the degree of timing tendency vary with the category recorded, and, if so, what are the possible causes of the variation?

(2) What is the extent of timing tendency in the total record of behavior when category distinctions are disregarded?

(3) To what extent do individual observers vary in degree of

timing tendency within each category and in the total record when category distinctions are disregarded?

(4) Does the size of the timing tendency index decrease with repeated observations of the same behavior, i. e., do observers improve with practice in observing the same events several times?

(5) To what extent does the size of index vary with different films? Is there any evidence of improvement with repeated observation of the same types of behavior in different films?

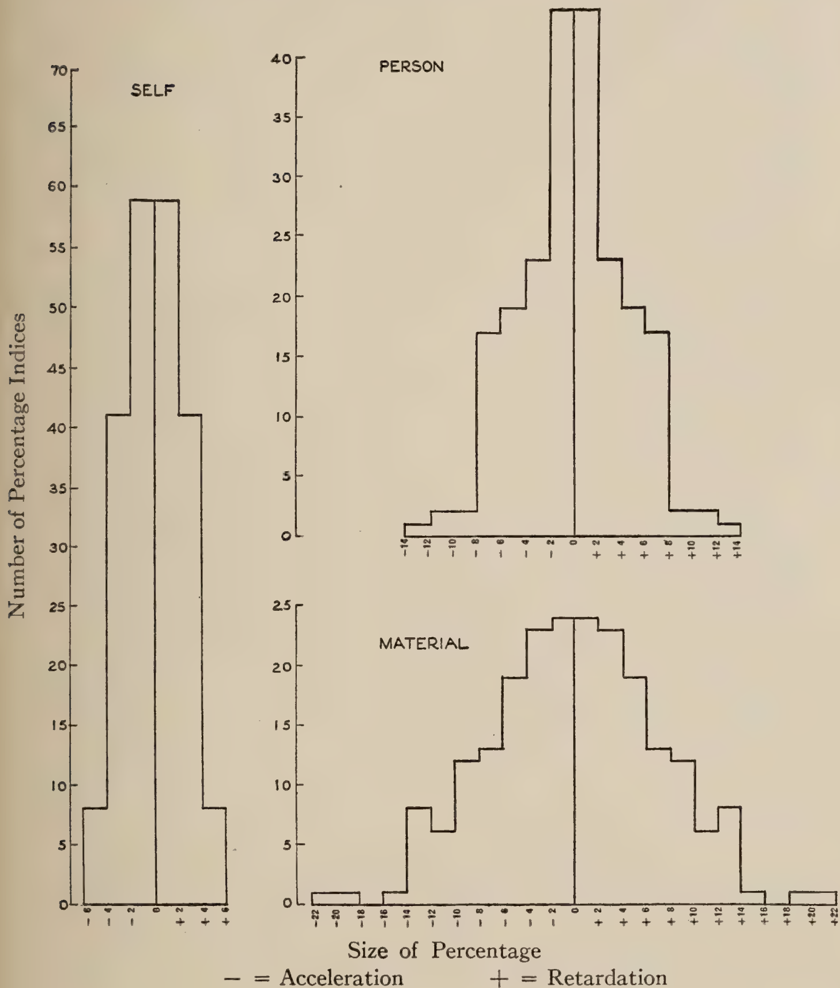
Individual patterns among observers with regard to the direction of timing tendency, i. e., consistent biases toward retardation or acceleration of records, will be discussed in connection with other forms of bias in Chapter IX.

Extent of Timing Tendency per Category of Behavior

Histogram Series I shows graphically the distribution of the 216 pairwise indices of timing tendency per category which were shown in Table XI. Since, in our method of analysis, the accelerated tendency of one observer in a given pair is exactly balanced by an equivalent retarded tendency of the other, and vice versa, these distributions which include indices for both members of each pair will of necessity cover exactly equivalent areas on either side of the 0 point. Comparing the histograms for *Material*, *Self* and *Person*, we find that percentages of timing tendency for the *Material* category cover the largest range (− 20.9% to + 20.9%), those for *Person*, the next largest (− 12.5% to + 12.5%) and those for *Self*, the smallest (− 4.9% to + 4.9%). One hundred and ten of the 216 timing indices for *Material* fall within the range − 5% to + 5%, 154 indices for *Person* and the entire 216 for *Self*. Seventy-two indices for *Material* and 56 for *Person* fall between − 5% and − 10% or between + 5% and + 10%. Thirty-four indices for *Material* and six for *Person* are as large as, or larger than, 10%. It appears, therefore, that all observers tend to have larger percentages of timing tendency for *Material* than for either of the other categories and larger percentages for *Person* than for *Self*. The same relationship appears with uniform consistency in the average percentages of timing tendency for each observer paired with three other observers in all observations of Series A (Table XII).

The complete consistency of this relationship suggests that other factors, in addition to the nature of the behavior observed,

HISTOGRAM SERIES I. DISTRIBUTION OF 216 PAIRWISE PERCENTAGE INDICES OF TIMING TENDENCY BY CATEGORY, SERIES A.



may be operating to produce this variation in size of percentage for different categories. Since the incidence of the three types of behavior was known to have differed markedly, the group

percentages of timing tendency per category for each of the six films of Series A (three observations per film totaled), for the total of Series A (all films and observations totaled), and for the third observation of Series C, normal and slow speed, were compared with the bases on which the percentages were computed, i. e., with the average number of weighted "synchronous" units per observer. The following table summarizes the data on this point.

Observation Series	Film	Material		Self		Person	
		Base	%	Base	%	Base	%
A	I	1316	12.0	8893	4.0	2368	7.6
	II	1771	5.4	14871	2.7	5198	5.5
	III	665	10.5	10703	2.4	1667	2.3
	IV	1590	9.3	17540	2.3	4278	4.4
	V	1978	5.1	12778	3.6	1541	5.7
	VI	2772	5.0	14817	2.3	3704	4.6
	I-VI	10092	6.7	79602	2.8	18756	4.9
C	Normal Speed	1124	1.4	773	4.5	314	7.0
	Slow Speed	3519	2.5	2938	3.2	816	8.8

These figures indicate that infrequency of occurrence of a given sort of behavior is itself associated with irregularities in timing this behavior. Thus, the lower percentages of timing discrepancy seem to occur, irrespective of category, in direct relation to the proportionate frequency of the category. For the whole of Series A, the lowest timing tendency index is found for *Self*, the next lowest for *Person*, and the highest for *Material*. The frequency of occurrence of behavior is highest in *Self*, next highest in *Person* and lowest in *Material*. In the six separate films of Series A, the relationship is consistent with the findings for the total of the Series. The lowest indices, with one exception (Film III), are in the *Self* category, and *Self* is consistently the most frequently recorded behavior. *Material* is less frequently recorded than *Person* in five of the six films and, in four out of these five, the index of timing tendency for *Material* is higher than that for *Person*, whereas in the film where *Person* is less frequently recorded (Film V) the index for *Person* is higher than that for *Material*. For Series C, the lowest indices of timing tendency are found for *Material*, the next lowest for *Self* and the

highest for *Person* and here we find the bases varying from low to high, respectively, from *Person* to *Self* to *Material*.

There are three possible factors involved in these differences,—(1) the mere fact of a small base in itself increases the degree of uncertainty of percentages computed upon it, (2) infrequency of occurrence probably increases observational difficulties and makes such categories, irrespective of the nature of the behavior observed, more liable to erroneous recording, and (3) there are undoubtedly real differences in the difficulty of identifying the three categories of behavior which are reflected in these timing discrepancies. It is impossible completely to disentangle these three factors. Differential interpretation difficulties will be considered in the next chapter. The other two factors may be partially disentangled by computing the standard deviation of each of the timing tendency indices,[1] considering the probable degree of uncertainty to be three times this standard deviation and determining whether there is any real difference between the indices when due allowance has been made for these varying degrees of uncertainty. This analysis is made with due caution and recognition of the possible inapplicability of concepts based on random sampling.

The following table shows the differences between the timing tendency indices of each pair of categories for each film of Series A, for the total of Series A, and for the third observation of Series C, normal and slow speed,[2] together with the standard deviations of these differences multiplied by 3, these latter figures being taken as indicating the probable degree of uncertainty of the differences. If the differences are not consistently greater than three times their standard deviations, we may infer that a major contribution to the differences was the mere random fluctuation due to sampling and largely attributable to a small base as such. If the differences are consistently greater than three times their standard deviations, we may assume that other factors played an important rôle. For the total of Series A, the differences between the indices seem to be of such a degree of significance as to indicate that they cannot be attributed merely

[1] See Appendix E for reference tables giving the standard deviations of percentages from 0.5 to 20.0 computed on different bases.

[2] See table, p. 136.

to the differences in the size of the base used in computing these indices, but that the rank from high to low in the *Material*, *Person* and *Self* categories must be attributed to other factors, of which, however, infrequency of occurrence may very probably be one.[1]

Paired Categories		Series A Films							Series C	
		I	II	III	IV	V	VI	Total	Normal	Slow
Material-Self	Diff.	+8.0	+2.7	+8.1	+7.0	+1.5	+2.7	+3.9	−3.1	−0.7
	3σ Diff.	2.8	1.5	3.6	2.1	1.2	1.2	0.8	2.5	1.3
Material-Person	Diff.	+4.4	−0.1	+8.2	+4.9	−0.6	+0.4	+1.8	−5.6	−6.3
	3σ Dif.	3.1	1.7	3.7	2.3	1.9	1.5	0.9	4.4	3.1
Self-Person	Diff.	−3.6	−2.8	+0.1	−2.1	−2.1	−2.3	−2.1	−2.5	−5.6
	3σ Diff.	1.6	1.0	1.0	1.0	1.5	1.0	0.5	4.9	3.1

That this is the case is suggested by Series C, where the small base, as such, apparently accounted for a large part of the differences in the indices, but where the high indices for *Person* and the low indices for *Material* show somewhat greater than expected differences in *Material-Self* and *Material-Person* (normal series) and in *Material-Person* and *Self-Person* (slow series). Looking at Series A in detail, we find that the differences between the indices for *Material* and for Self are apparently accounted for in no instance by the small base alone, since for each of the six films the actual difference between the paired indices is more than three times as great as the standard deviation of the difference. The differences between the indices for *Self* and *Person* are, with one exception, Film III, greater than the expected differences. In the *Material-Person* comparison, the differences apparently were accounted for by the small base in three of the six films. This suggests that there are important differences in the *Self* category, as compared with *Material* and *Person*, which are productive of much less timing discrepancy. That mere difference in the size of base is not the sole determining factor seems clear, but it is probable that infrequency of the behavior in a given category is a factor in producing timing discrepancy.

For the third observation of each film of Series A, the average base per observer and the group percentage derived from it are shown below.

[1] The formula $\sqrt{(\sigma_{P_1})^2 + (\sigma_{P_2})^2}$ was used for determining the standard deviations of the differences between percentages.

Film	Material		Self		Person	
	Base	%	Base	%	Base	%
I	460	12.8	3091	3.7	809	10.5
II	643	4.0	4989	2.5	1770	4.9
III	237	10.1	3604	1.9	537	2.2
IV	559	5.7	5899	1.7	1422	3.9
V	733	2.4	4381	3.6	524	6.5
VI	1039	5.2	4985	1.9	1261	4.3

A comparison of paired categories similar to that just given for the total of the three observations of each film follows.

Paired Categories		Films					
		I	II	III	IV	V	VI
Material-Self	Diff.	+9.1	+1.5	+8.2	+4.0	−1.2	+3.3
	3 σ Diff.	4.8	2.4	5.9	3.0	1.9	2.1
Material-Person	Diff.	+2.3	−0.9	+7.9	+1.8	−4.1	+0.9
	3 σ Dif.	5.7	2.8	6.2	3.3	3.7	2.7
Self-Person	Diff.	−6.8	−2.4	−0.3	−2.2	−2.9	−2.4
	3 σ Diff.	3.4	1.7	2.0	1.6	3.3	1.8

Here the differences between the indices for *Material* and for *Self* are apparently accounted for by the small base alone in only two of the six films and the same is true of *Self-Person* but the comparison for *Material-Person* shows no significant degree of difference in four of the six films. Five of the eight differences between paired categories which here are smaller than expected become greater than the expected differences in the comparison based on the larger amount of data (three observations per film combined) represented by the table on p. 138; the other three become less significant. Of the differences that here are greater than three times their standard deviations, i. e., in the data for the third observation only, eight become more significant and two less significant in the data for the three observations combined. It seems obvious, therefore, that real differences between categories which are obscured in the data for a single observation, become clear-cut when the amount of data is increased.

Another factor affecting the extent of timing tendency is the number of breaks in the continuity of behavior within a category. Any change in category of behavior which lasted at least one unit[1] provided an opportunity for timing disagreement, since

[1] Breaks of less than one unit, i. e., less than 2½″, were not considered in this study of timing tendencies, since they were disregarded in our method of tabulation of the data. See Appendix A.

one observer might record the change as occurring in an earlier or later unit than another. The number of breaks, so defined, is computed by the simple process of adding the beginnings and endings of all blocks of behavior in each category, a block being defined as a continuous record of behavior within a category. The number of breaks per 100 units, by this definition, would approach 200 as an upper limit, this upper limit being reached only if each block of behavior is exactly one unit in length. The more frequent the breaks in a record of material, self or person behavior, i. e., the less continuous the record, the higher is the proportion of timing discrepancy to be expected. We would expect, in other words, a positive correlation between the average number of breaks in the continuity of behavior per category and the extent of timing tendency per category.

The following table, giving the number of breaks, as defined above, per 100 units of behavior for the third observation of the six films of Series A for each observer, together with the average for the group of four observers, shows the same relationship between the three categories as that previously noted. The number of breaks is largest for the *Material* category, next largest for *Person* and smallest for *Self*. The more broken the record of behavior in a given category, therefore, the greater is the chance for timing discrepancy and the larger the percentage of retardation or acceleration tendency.

Observer	Material						Self						Person					
	Films						Films						Films					
	I	II	III	IV	V	VI	I	II	III	IV	V	VI	I	II	III	IV	V	VI
W	87.1	60.3	92.7	76.5	83.8	68.3	29.3	23.6	22.8	19.5	28.4	28.4	64.0	38.9	32.6	33.2	59.6	59.6
X	86.0	61.5	81.0	72.0	78.9	65.6	26.4	22.4	21.7	20.0	28.9	27.5	65.7	39.3	39.1	29.2	49.0	51.5
Y	63.5	50.0	74.5	67.3	81.8	61.0	28.3	21.8	22.2	19.7	28.9	26.9	49.1	32.8	38.0	33.5	50.0	52.9
Z	78.2	45.5	77.3	71.3	83.4	74.3	26.6	25.0	26.8	23.4	35.3	29.4	51.9	38.4	35.4	32.7	50.5	45.4
Average	77.9	54.1	80.9	71.7	82.0	67.3	27.6	23.2	23.3	20.6	30.3	28.0	57.2	37.3	36.3	32.1	52.2	52.1

The differences noted seem to hold also with a fair degree of consistency in terms of the separate films for a given category. Because of the unreliability of small differences between indices, a rank comparison is inappropriate. Considering, however, the two highest and the two lowest indices for each category, for the

third observation of Series A, we find that the average breaks per observer in the *Material* category are 77.9 and 80.9 per 100 units of behavior for the films having the highest indices of timing discrepancy and 82.0 and 54.1 for those having the lowest indices; 27.6 and 30.3 for the highest indices in the *Self* category against 20.6, 23.3 and 28.0 for the three lowest; and 57.2 and 52.2 for the highest indices in the *Person* category against 36.3 and 32.1 for the lowest.

Extent of Timing Tendency in Total Record of Behavior

What is the extent of timing tendency when the total record of behavior is considered, regardless of interpretation, i. e., when merely the timing of the presence of a character on the motion picture screen is involved? To what extent do the factors we have noted in connection with the separate categories affect also the percentages of timing tendency based on the total record?

It will be noted that the percentage indices, as shown in Table XIII, are uniformly smaller than those for the separate categories (Tables XI, XII). None exceeds 3.2% and, for the pairs WX, WY and XY, the highest percentage is 1.8%. It is clear that the size of these indices is necessarily affected by the same factors which have been discussed in connection with the separate categories, *Material*, *Self* and *Person:* namely, by the size of base on which percentages were computed and by the number of breaks in the continuity of record. The bases of the percentages for the total record regardless of category are, of course, larger in each instance than those for separate categories, but are not equal to the sum of the totals of the three categories because of our practice of considering fractions of a unit equivalent to a whole unit in the category analysis.

In addition to the larger number of recorded units of behavior in the total record, the greater continuity of the record, when categories are disregarded, operates to reduce the size of percentages of timing tendency. Breaks within a category, as has been indicated, represent changes in type of behavior while the character is on the screen. Breaks in the total record without regard to category occur only when the character appears upon the screen or disappears from it, i. e., they represent entrances and exits rather than changes in behavior. The number of

breaks (lasting at least one unit) in continuity of behavior per 100 units of record for each observer's total record of behavior regardless of category and the average number for the four observers, in the third observation of Films I–VI, Series A, are shown in the following table. These figures are obviously smaller than those for any category considered separately. With the

Observer	Films					
	I	II	III	IV	V	VI
W	15.7	11.8	17.7	11.5	21.1	16.1
X	12.9	11.7	17.3	11.6	20.5	15.7
Y	14.1	11.4	16.6	11.8	20.8	16.2
Z	14.3	10.9	20.0	14.2	25.7	15.8
Average	14.3	11.4	17.9	12.2	22.0	16.0

exception of Film V for all observers and Film III for Observer Z, the breaks average less than 20 per 100 units of record. There are, consequently, fewer chances for timing errors and the percentages of retardation or acceleration are necessarily lower. Since both the factors mentioned tend to reduce the size of timing tendency index for the total record, we cannot interpret the figures as indicating greater accuracy on the part of observers in timing entrances and exits of the character observed than in timing changes of behavior from one category to another within the same scene, though this assumption is plausible. In the film in which the average number of breaks per observer is smallest (Film II), the percentage indices are, disregarding direction of the tendency, 0, 0.7, 0.9 and 1.6 for W, X, Y and Z, respectively; and in Film V, where breaks occur most frequently for all observers, the percentages are 0.4, 1.5, 0.8 and 2.7 for W, X, Y and Z. Thus there seems to be a definite tendency for frequency of breaks in continuity of behavior to be associated with high indices of timing tendency, and for infrequency of breaks to be related to low indices.

Extent of Timing Tendency per Observer

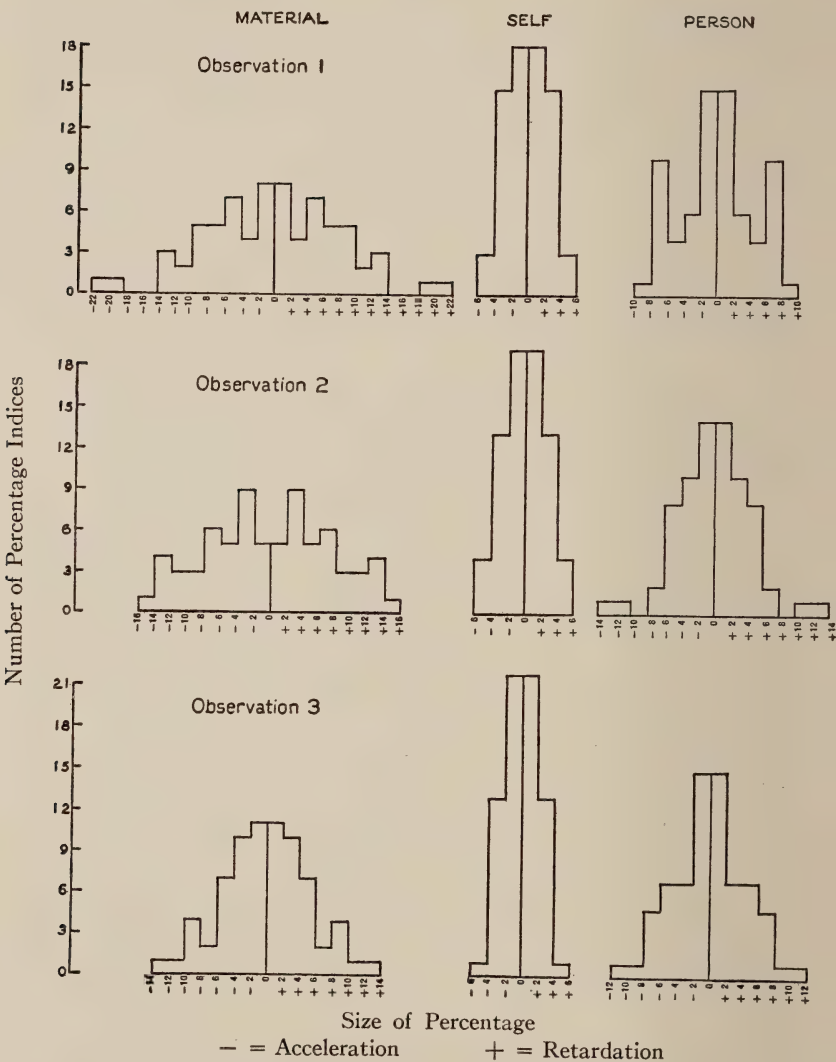
We have noted a uniform difference in size of timing tendency index for different categories for all observers. To what extent do individual observers vary within each category in direction

and extent of timing tendency? The distribution of timing indices per observer and category for all observations of Series A (i. e., the 54 pairwise indices for each observer and category shown in Table XI) is graphically represented in Histogram Series II. Observer Z has a pattern, with respect to direction of timing tendency, which is distinctly different from those of the other three observers, having only two retarded indices in *Material*, none in *Self*, and one and a half[1] in *Person*. Observers X and Y have similar patterns with regard to direction of timing tendency, with three-fourths and two-thirds, respectively, of their indices for the three categories on the + side of the o point, i. e., retarded. W approaches most nearly an even balance between acceleration and retardation, having slightly more than half of his indices retarded.

It is obvious, from the indices shown in Table XI, that the size of index varies considerably with the pairing of observers, the largest indices occurring consistently in the pairs in which Z is included. The highest percentage for *Material* is that for the combination XZ in the first observation of Film III (20.9%); the next highest, for the pair WZ in the first observation of Film I (19.5%); the next, for YZ in the second observation of Film IV (15.5%). In 18 observations of *Material*, the pairs XY and WX have only four indices each and WY only five indices which exceed 5.0%, whereas the pairs WZ and XZ have 12 indices each, and YZ, 15 indices greater than 5.0%. In the *Self* category, 2.4% is the largest index for any pair which does not include Z, while the pairs WZ, YZ and XZ have 12, 13 and 14 indices, respectively, that are as high as or higher than 2.4%. In *Person*, WZ, XZ and YZ have six, ten and seven indices, respectively, larger than 5.0%, as compared with three for WX and two each for WY and XY. In the indices based on the total record of behavior (Table XIII) we find the same situation. The percentages for Observers W, X and Y when paired with Z are consistently higher, with the exception of one pairing, than are their percentages when paired with one another. Obviously, the extreme deviation of Z from the rest of the group increases the size of the percentages for all observers paired with him. Since behavior cannot be recorded before it occurs, our inference must

[1] Zero indices were divided equally between the intervals on either side of the o point.

Histogram Series II. Distribution of 54 Pairwise Percentage Indices of Timing Tendency per Observer and Category, Series A.



be that Z either habitually makes his record in earlier time intervals than other observers, i. e., has a recording bias, or that his response to the observed events is quicker and therefore more

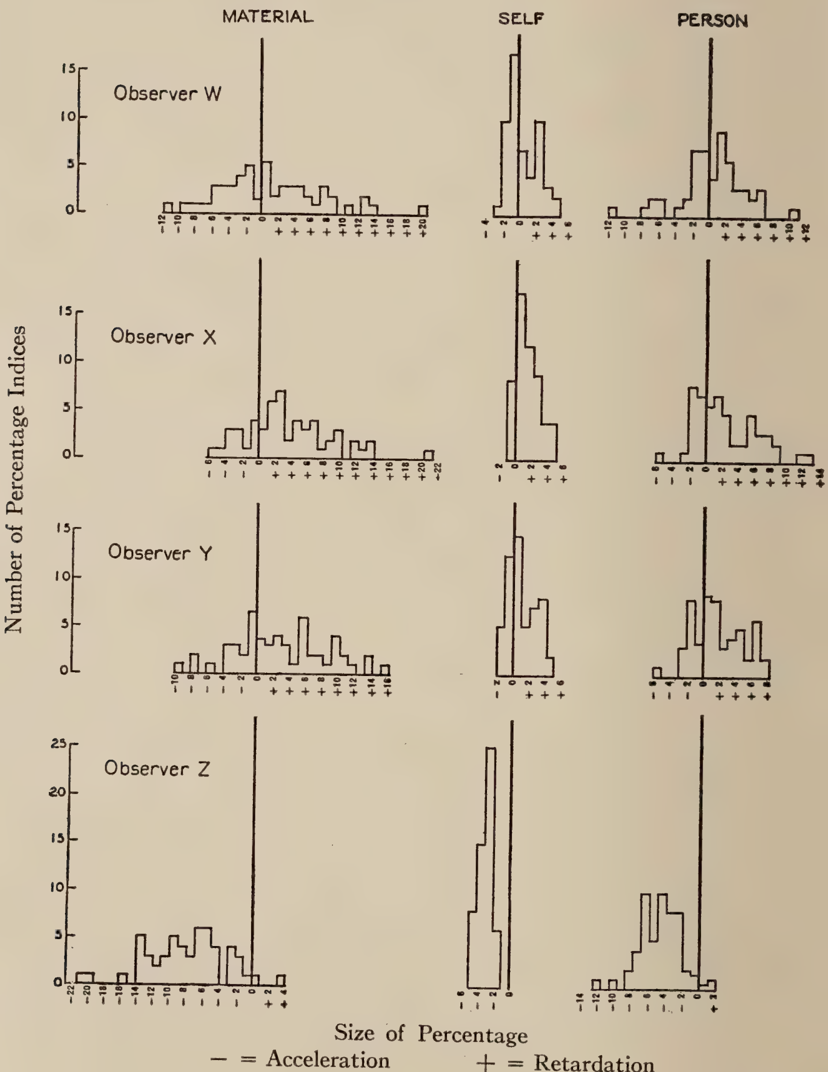
accurate than that of his co-observers. The tendencies of the other observers toward retardation make them appear to be more reliable and the fact that all are retarded with regard to Z makes any pair in which Z appears show excessive timing discrepancy. Since our method of comparison is entirely relative, we can only infer from these findings that Z's record of behavior is probably more accurately timed than those of the other observers.

Improvement with Repeated Observation of the Same Behavior

To what extent does the size of timing tendency index diminish with repeated observation of the same behavior? Are observers consistent in consecutive observations, or do they show improvement, i. e., become more reliable, as indicated by progressively lesser degrees of timing tendency when they observe the same series of situations three successive times? Such improvement would be indicated by a decrease in size of index, a closer approximation to a 0 balance between acceleration and retardation. An attempt to indicate possible improvement is made in three ways,—(1) by presenting histograms showing the clustering of the indices of timing tendency around 0 for each of the three consecutive observations, by categories, (2) by examining the separate indices to see to what extent there was a change toward less timing tendency in consecutive observations, irrespective of amount, and (3) by estimating whether the *amount* of apparent improvement in consecutive observations was reliable or was due to fluctuations of sampling. We feel that no one of these analyses is, in itself, adequate, but the three, taken together, will be indicative of the fact of improvement even if they fail in estimating the degree.

In Histogram Series III the indices shown in Table XI are grouped by category and observation. There appears to be relatively little improvement in *Material* between the first and second observations of the six films of Series A, since, with the exception of the six extreme indices, the two distributions cover practically the same range (− 14% to + 14%). Improvement is indicated, however, by the fact that the highest point of the distribution for Observation 2 is lower than that for Observation 1. A more marked improvement is shown in the third observa-

Histogram Series III. Distribution of 72 Pairwise Percentage Indices of Timing Tendency by Category and Observation, Series A.



tion of *Material*. Forty-six indices, as compared with 32 each for the first and second observations, are concentrated within

the range − 5% to + 5%, 22 fall between − 5% and − 10% or between + 5% and + 10%, and only 4 exceed 10%, as compared with 14 for the first and 16 for the second observation. In records of *Self*, the range is the same for first, second and third observations, but there is a tendency toward increase in proportion of indices clustered about the center of the distribution as we compare histograms for the first and second observations, and for the second and third. In records of *Person*, improvement is slight and not uniform for all pairs of observers. Indices for the second and third observations in this category cover a wider range than those for the first, indicating that there was less timing discrepancy in the first observation than in the later ones.

Looking now at the direction of the change in the indices, irrespective of the amount of change, we may expect a consistent decrease in size of index if there is a consistent tendency toward improvement. Are the indices for the same films consistently nearer 0 in the second than in the third observation, in the third than in the second, and in the third than in the first, observation?

Each observer has seen six films with each of three other observers and therefore has 18 chances of improvement from the first to the second observation, from the second to the third, and from the first to the third. It is clear from the following table that improvement in timing, as indicated by a regressive tend-

Category	Observer	Number of Chances for Improvement	Tendency on Second Observation nearer 0 than on First	Tendency on Third Observation nearer 0 than on Second	Tendency on Third Observation nearer 0 than on First
Material	W	18	8	11	10
	X	18	9	8	11
	Y	18	7	12	10
	Z	18	8	11	15
	All	72	32	42	46
Self	W	18	11	9	15
	X	18	8	12	11
	Y	18	11	12	12
	Z	18	12	15	16
	All	72	42	48	54
Person	W	18	9	9	13
	X	18	11	8	11
	Y	18	11	8	13
	Z	18	11	7	13
	All	72	42	32	50

ency toward 0 on consecutive observations, can be said to occur in a clear-cut manner in the *Self* category where, on the average, 58% (42/72) of the indices for the second observations are regressive as compared with the first, two-thirds (48/72) of the indices for the third observations are regressive as compared with the second, and three-fourths (54/72) are regressive in the third as compared with the first. The *Material* category shows this regressive tendency in 44% (32/72) of the indices for second observations compared with first, in 58% (42/72) of the indices for the third compared with the second and in 64% (46/72) for the third compared with the first. The *Person* category shows the greatest regressive tendency in the third observations compared with the first, 69% (50/72) of the indices being regressive. Fifty-eight per cent (42/72) of the indices for the second observations compared with the first are regressive and only 44% (32/72) of those for the third compared with the second. There is apparently no generally consistent tendency toward improvement as manifested by regressive tendencies in all indices and categories, but it occurs so frequently that significant differences will probably be shown in the average tendencies.

Because of the difference in size of the bases used in computing these percentage indices of timing tendency, the frequency distributions can be taken as only grossly representative of any real tendency toward improvement. Table XII presents a clearer picture of the general tendencies. For the six films combined, there is evidence of progressive improvement in timing in the group figures for all three categories, from the first to the second, and from the second to the third, observation. The standard deviation of the average tendency per observation, for six films combined, is for *Material*, .312, for *Person*, .194, and for *Self*, .072. We may, therefore, assume that there is clear evidence of improvement if the differences in the index for *Self* are of the order of 0.3, for *Person*, of 0.8, and for *Material*, of 1.3. This estimate was obtained by using a modification of the usual formula—three times the square root of the sum of the squares of the standard deviations—as the boundary for determining whether the differences are due to "chance" or are "significant." Since the standard deviations of the percentage tendencies on consecutive observations were here approximately the same, we

have simply computed σ for the average tendency in relation to the average base and taken the boundary of significance as being $3\sqrt{2(\sigma^2)}$, an estimate considered to be sufficiently accurate for our purposes. On this basis, the apparent improvement between the first and second observations in *Material* is probably due to chance fluctuations, whereas the improvement between the first and third, and between the second and third, observations is more likely to be a true improvement. The improvements noted in the *Self* category between the first and second, second and third, and first and third, observations are all probably reliable. In *Person*, the improvements between the first and second, and the first and third, observations are probably real, but the apparent slight improvement between the second and third is unreliable.

For judging the differences between the separate films, the following estimates of "real" improvement, computed in the same manner, may be accepted for the group: in *Material*, differences of the order of 5.1 for Film I, 3.0 for Films II and V, 7.2 for Film III, 3.8 for Film IV and 2.1 for Film VI; in *Self*, differences of the order of 0.8 for Films I, III and V, and of 0.4 for Films II, IV and VI; and in *Person*, differences of 3.0 for Film I, 1.7 for Films II, IV and VI, 2.1 for Film III and 3.4 for Film V. On the basis of these estimates, genuine improvement may be said to occur for the group in Films II and IV, all categories, in Films III and V, *Material*, and in Film VI, *Self* and *Person*. There is no evidence of "real" improvement in any category in Film I.

Improvement with Observation of Similar Behavior in Different Films

We have noted a general tendency for observers to improve in timing behavior in successive observations of the same events. Do they also improve as a result of practice in observing the same types of behavior in different situations over a period of time? Does the size of timing tendency index decrease progressively from film to film as well as from observation to observation of the same film, or do individual differences in films obscure any such trend toward general improvement? The latter seems to be the fact, whatever measure of differentiation we apply. No

generally consistent decrease in size of index can be found either in the percentages based on uncombined data per pair (Table XI) nor in the figures based on combined data for three pairings of each observer for successive films of Series A (Table XII). The group indices per film and category show excessive variation in certain categories for certain films, but no consistent decrease within each category from the first to the sixth film (Table XII). Film I has the largest index in all three categories. Films III and IV have high indices for *Material*, Film V for *Self*, and Films II and V for *Person*. The corresponding group figures for total record of behavior regardless of interpretation correspond in relative size per film rather closely to those for the *Self* category. Films II, III, IV and VI (total record) have highly similar indices, ranging from 1.5 to 1.8%. Films I and V have group indices of 2.6% and 2.7% respectively (Table XIII). We may conclude, therefore, that the behavior of characters in different motion picture films varies in observability to such an extent that no general improvement in timing reliability can be discovered in tendencies for successive films. This conclusion should be referred to data on the variations in interpretation indices from one situation to another in Chapter X.

Amount of Timing Discrepancy

Our discussion thus far has been concerned with the direction and extent of timing tendencies of different observers. A method of measuring the excess of retarded over accelerated, or of accelerated over retarded, records has been described, and indices of timing tendency for four observers, computed by this method, have been discussed. Two observers, we find, tend to retard more records than they accelerate, one shows the reverse tendency, and one shows the expected tendency, an approximate balance between retardation and acceleration. What proportion of an observer's total recorded units is non-synchronous with the corresponding units of other observers but adjacent to such units? What percentage is retarded, what percentage accelerated, and what is the total proportion of timing discrepancy without regard to direction of the tendency? In measuring this gross amount of timing error, the same major assumption as to probability of occurrence was made as in the analysis of direc-

tion of timing tendency: namely, that the greater the number of persons recording an event in a given time unit, the greater is the likelihood that the event occurred in that unit. Only single units preceded or followed by units of both observers were counted in this analysis. In pairwise comparison of observers, every discrepant unit of an observer preceded by units by that observer and by the paired observer was considered retarded, and every discrepant unit followed by units by both members of the pair was counted as an acceleration. The number of retarded units and the number of accelerated units, thus determined, for each observer paired with each other observer, were in turn divided by the total number of units recorded by the given observer, to obtain the proportion of units retarded and accelerated. Since the retarded and accelerated units measured in this way were mutually exclusive, the sum of the two, divided by the total number of units, gives the total per cent of divergently timed units. These percentages per film, observation, category and observer are given in Table XIV. The per cent of discrepantly timed units of *Material*, for the 18 observations of Series A combined, ranges from 13.1 (Observer X) to 17.4 (Observer Y); for *Self*, from 6.2 (Observer Z) to 7.1 (Observer X); and for *Person*, from 8.0 (Observer W) to 12.0 (Observer Z). Percentages per film for different observers range from 11.1 to 21.4 for *Material;* from 4.4 to 8.9 for *Self;* and from 5.4 to 17.6 for *Person*.

Improvement of observers with consecutive observation of the same behavior, as indicated by decrease in the size of percentages of discrepantly timed units from the first to the third observation (all films combined) without regard to retarded or accelerated tendency, is slight, and not uniform for all observers and all categories. (Table XIV). The arithmetic average of the 12 percentages of total timing error for the six films combined (four observers, three observations) is 15.0 for *Material*, 6.7 for *Self*, and 10.0 for *Person*. (Column I–VI, Table XIV.) Using the same method for estimating "real" improvement as that used for the data on timing tendencies, we find that a difference between percentages in the *Material* category should be of the order of 3.4 to be significant; in the *Self* category, a difference of 0.8 suggests that the improvement is genuine; and in the *Person* category, a difference of 2.1. On this basis, none of the four

TABLE XIV

Percentages of Total Recorded Units Retarded and Accelerated, and Percentages of Total Units Discrepantly Timed, without regard to Direction of Timing Error, for Each Observer (all pairings combined) by Film, Observation and Category of Behavior (Series A)

Standard Observer	Co-observers	Direction of Timing Error	Observation	Material							Self							Person						
				Films							Films							Films						
				I	II	III	IV	V	VI	I–VI	I	II	III	IV	V	VI	I–VI	I	II	III	IV	V	VI	I–VI
W	XYZ	Retarded	1	10.4	10.3	9.3	6.2	8.5	8.2	8.7	5.9	4.2	2.9	2.9	5.0	4.0	4.0	5.6	3.6	3.3	3.4	5.3	3.2	3.9
			2	6.9	7.2	8.9	7.2	5.0	9.3	7.6	5.8	3.5	3.7	3.0	4.1	4.0	3.9	7.5	4.4	3.7	3.2	5.5	4.4	4.6
			3	6.5	7.4	9.7	4.5	6.3	9.0	7.2	3.3	3.2	3.7	3.4	4.3	3.5	3.6	4.8	5.3	4.4	4.0	5.0	4.2	4.6
			Av.	7.7	8.3	9.3	6.0	6.6	8.9	7.8	5.0	3.6	3.4	3.1	4.4	3.9	3.8	6.0	4.5	3.8	3.5	5.2	4.0	4.4
		Accelerated	1	7.3	6.5	9.2	5.4	6.7	8.8	7.2	4.1	3.8	2.0	2.7	4.5	3.9	3.5	4.8	2.8	2.2	2.7	5.3	4.8	3.5
			2	4.4	5.8	4.3	9.1	5.1	7.2	6.3	3.1	3.0	2.3	2.5	3.9	2.9	2.9	4.9	4.2	1.0	2.5	5.4	4.9	3.9
			3	6.1	6.9	4.9	5.4	8.2	4.4	6.0	3.6	2.7	2.8	2.6	3.7	3.0	3.0	7.8	2.4	1.8	1.7	6.0	3.4	3.5
			Av.	5.9	6.4	6.0	6.7	6.7	6.6	6.4	3.6	3.2	2.4	2.6	4.1	3.2	3.2	5.9	3.1	1.6	2.3	5.6	4.3	3.6
		Total	1	17.7	16.8	18.5	11.6	15.2	17.0	15.9	10.0	8.0	4.9	5.6	9.5	7.9	7.5	10.4	6.4	5.5	6.1	10.6	8.0	7.4
			2	11.3	13.0	13.2	16.3	10.1	16.5	13.9	8.9	6.5	6.0	5.5	8.0	6.9	6.8	12.4	8.6	4.7	5.7	10.9	9.3	8.5
			3	12.6	14.3	14.6	9.9	14.5	13.4	13.2	6.9	5.9	6.5	6.0	8.0	6.5	6.6	12.6	7.7	6.2	5.7	11.0	7.6	8.1
			Av.	13.6	14.7	15.3	12.7	13.3	15.5	14.2	8.6	6.8	5.8	5.7	8.5	7.1	7.0	11.9	7.6	5.4	5.8	10.8	8.3	8.0
X	WYZ	Retarded	1	10.6	6.0	10.6	9.1	9.3	4.8	7.8	5.2	4.3	3.5	4.2	6.4	5.0	4.7	7.6	4.6	1.7	5.0	7.1	5.4	5.2
			2	6.4	5.7	10.7	10.4	7.8	6.2	7.4	5.7	3.5	3.3	4.1	5.5	4.2	4.3	7.3	5.0	2.6	3.4	7.5	7.8	5.5
			3	9.3	6.2	13.5	10.5	8.7	7.6	8.6	5.3	3.1	4.1	3.2	5.7	3.4	4.0	9.1	4.4	3.6	5.4	8.2	5.5	5.7
			Av.	8.8	6.0	11.6	10.0	8.6	6.2	7.9	5.4	3.6	3.6	3.9	5.9	4.2	4.4	8.1	4.7	2.6	4.6	7.6	6.2	5.4
		Accelerated	1	5.6	7.2	4.0	4.3	4.6	5.3	5.2	4.0	3.0	2.6	2.2	3.3	3.3	3.0	3.8	3.4	3.7	2.4	3.6	5.2	3.6
			2	3.0	1.5	4.7	5.9	4.5	5.6	4.4	3.3	2.2	2.1	2.3	3.1	3.0	2.6	3.6	3.2	3.2	2.8	4.4	3.9	3.4
			3	5.0	6.9	3.2	5.2	7.0	5.8	5.9	3.3	2.6	2.8	1.9	2.6	2.9	2.6	3.2	2.9	1.5	2.5	3.0	4.4	3.1
			Av.	4.5	5.1	4.0	5.1	5.5	5.6	5.2	3.5	2.6	2.5	2.1	2.9	3.1	2.7	3.5	3.2	2.9	2.6	3.7	4.5	3.4
		Total	1	16.2	13.2	14.6	13.4	13.9	10.1	13.0	9.2	7.3	6.1	6.4	9.7	8.3	7.7	11.4	8.0	5.4	7.4	10.7	10.6	8.8
			2	9.4	7.2	15.4	16.3	12.3	11.8	11.8	9.0	5.7	5.4	6.4	8.6	7.2	6.9	10.9	8.2	5.8	6.2	11.9	11.7	8.9
			3	14.3	13.1	16.7	15.7	15.7	13.4	14.5	8.6	5.7	6.9	5.1	8.3	6.3	6.6	12.3	7.3	5.1	7.9	11.2	9.9	8.8
			Av.	13.3	11.1	15.6	15.1	14.1	11.8	13.1	8.9	6.2	6.1	6.0	8.8	7.3	7.1	11.6	7.9	5.5	7.2	11.3	10.7	8.8

TABLE XIV—*Continued*

Percentages of Total Recorded Units Retarded and Accelerated, and Percentages of Total Units Discrepantly Timed, without regard to Direction of Timing Error, for Each Observer (all pairings combined) by Film, Observation and Category of Behavior (Series A)

Standard Observer	Co-observers	Direction of Timing Error	Observation	Material							Self							Person						
				Films							Films							Films						
				I	II	III	IV	V	VI	I–VI	I	II	III	IV	V	VI	I–VI	I	II	III	IV	V	VI	I–VI
Y	WXZ	Retarded	1	11.0	9.5	10.4	9.6	11.1	10.2	10.3	4.7	4.1	3.9	3.5	5.1	5.1	4.4	10.2	7.9	4.4	6.9	6.2	8.2	7.6
			2	12.3	8.4	8.5	15.5	10.9	7.8	10.4	3.8	4.0	3.5	3.0	4.6	4.0	3.8	9.2	6.6	3.6	5.0	8.5	5.8	6.4
			3	13.8	10.8	9.5	9.2	9.1	8.8	10.0	3.6	3.1	3.8	3.3	5.0	3.5	3.7	10.0	6.3	5.4	3.4	5.9	5.9	6.0
			Av.	12.3	9.6	9.4	11.5	10.3	8.7	10.2	4.0	3.7	3.7	3.3	4.9	4.2	4.0	9.8	7.0	4.4	5.1	6.9	6.6	6.6
		Accelerated	1	9.1	7.5	8.1	9.2	7.0	3.7	7.3	3.2	2.5	2.4	2.2	3.6	3.3	2.8	7.3	4.3	3.3	3.4	5.6	2.5	4.2
			2	10.8	4.6	9.2	7.2	6.1	6.2	7.0	3.3	2.5	2.6	1.9	3.9	3.8	2.9	8.8	4.0	3.3	2.8	5.8	5.6	4.8
			3	7.4	6.9	12.1	5.8	8.1	6.4	7.3	2.1	1.8	2.2	2.1	3.4	3.4	2.5	5.7	4.3	3.3	2.7	5.3	5.7	4.4
			Av.	9.1	6.3	9.9	7.4	7.1	5.7	7.2	2.9	2.3	2.4	2.0	3.6	3.5	2.7	7.2	4.2	3.3	3.0	5.6	4.6	4.5
		Total	1	20.1	17.0	18.5	18.8	18.1	13.9	17.6	7.9	6.6	6.3	5.7	8.7	8.4	7.2	17.5	12.2	7.7	10.3	11.8	10.7	11.8
			2	23.1	13.0	17.7	22.7	17.0	14.0	17.4	7.1	6.5	6.1	4.9	8.5	7.8	6.7	18.0	10.6	6.9	7.8	14.3	11.4	11.2
			3	21.2	17.7	21.6	15.0	17.2	15.2	17.3	5.7	4.9	6.0	5.4	8.4	6.9	6.2	15.7	10.6	8.7	6.1	11.2	11.6	10.4
			Av.	21.4	15.9	19.3	18.9	17.4	14.4	17.4	6.9	6.0	6.1	5.3	8.5	7.7	6.7	17.0	11.2	7.7	8.1	12.5	11.2	11.1
Z	WXY	Retarded	1	4.9	5.6	4.5	4.9	5.0	4.1	4.8	2.3	1.8	2.3	1.3	2.4	1.6	1.8	7.2	2.8	2.7	1.9	4.7	4.9	3.7
			2	3.1	3.7	4.1	2.9	4.8	5.7	4.4	2.5	1.3	2.0	1.4	1.7	1.8	1.7	4.1	4.3	4.2	3.0	5.4	4.4	4.1
			3	6.7	8.0	8.7	5.2	8.3	3.5	6.3	2.8	1.7	1.9	1.2	1.6	2.0	1.8	4.9	3.3	3.5	2.2	3.7	4.7	3.6
			Av.	4.9	5.8	5.8	4.4	6.1	4.5	5.2	2.5	1.6	2.1	1.3	1.9	1.8	1.8	5.4	3.5	3.5	2.3	4.6	4.7	3.8
		Accelerated	1	13.5	10.9	15.2	11.6	8.2	7.3	10.1	5.8	5.0	4.9	4.1	6.1	4.6	5.0	12.9	9.2	3.0	9.0	11.0	8.5	9.2
			2	13.4	10.7	13.0	13.6	8.3	8.3	10.4	6.5	3.4	4.2	3.3	5.1	3.8	4.2	12.9	8.7	5.2	5.3	9.1	7.5	8.1
			3	13.6	11.0	11.0	6.9	6.9	8.4	9.1	6.3	4.5	3.3	2.1	5.0	3.9	4.0	10.8	7.2	3.1	5.3	6.8	8.8	7.3
			Av.	13.5	10.8	13.1	10.5	7.8	8.0	9.8	6.2	4.3	4.1	3.1	5.4	4.1	4.4	12.2	8.4	3.8	6.6	9.0	8.2	8.2
		Total	1	18.4	16.5	19.7	16.5	13.2	11.4	14.9	8.1	6.8	7.2	5.4	8.5	6.2	6.8	20.1	12.0	5.7	10.9	15.7	13.4	12.9
			2	16.5	14.4	17.1	16.5	13.1	14.0	14.8	9.0	4.7	6.2	4.7	6.8	5.6	5.9	17.0	13.0	9.4	8.3	14.5	11.9	12.2
			3	20.3	19.0	19.7	12.1	15.2	11.9	15.4	9.1	6.2	5.2	3.3	6.6	5.9	5.8	15.7	10.5	6.6	7.5	10.5	13.5	10.9
			Av.	18.4	16.6	18.9	14.9	13.9	12.5	15.0	8.7	5.9	6.2	4.4	7.3	5.9	6.2	17.6	11.9	7.3	8.9	13.6	12.9	12.0

observers shows genuine improvement in the recording of *Material* or of *Person.* Observers W and Y show improvement between the first and third observations, Observers X and Z, between the first and second, and between the first and third, observations of *Self.*

The corresponding estimates of differences indicative of "real" improvement for the separate films are as follows: in *Material,* 9.8, 8.1, 14.0, 8.5, 7.6 and 6.4 for the six films, taken in consecutive order; in *Self,* 3.0 for Film I, 2.1 for Films II and VI, 2.5 for Films III and V and 1.7 for Film IV; in *Person,* 7.2, 4.2, 5.9, 4.2, 8.1 and 5.1 for the six films, taken consecutively. On the basis of these estimates, W shows genuine improvement in *Self,* between the first and third observations of Films I and II; Y, in *Person,* between first and third observations of Film IV; Z, in *Self,* between first and second observations of Film II and between the first and third observations of Film IV. On the whole, the proportion of timing error, as we have measured it, tends to remain relatively constant in the three consecutive observations of the same behavior.

To what extent do the percentages of discrepantly timed units per observer, shown in Table XIV, reveal differences among films? For *Material,* Observer W's range of percentages per film is from 12.7 to 15.5%, a difference of 2.8% between the film with least, and the one with most, timing error. Observer X has a wider range (4.5%), from 11.1 to 15.6%. Observer Y differs even more markedly in amount of timing discrepancy per film, having a range of 7.0%, from 14.4 to 21.4%. Observer Z varies almost as widely from film to film in his records of *Material,* having a range of 6.4%, from 12.5 to 18.9%. In *Self,* Observer W has a range of 2.9% between the film with least, and the one with most, timing error; Observer X, 2.9%, Observer Y, 3.2%, and Observer Z, 4.3%. In the *Person* category, the range for W is 6.5%, for X, 6.1%, for Y, 9.3%, and for Z, 10.3%. If, as a crude method of estimating the relative amount of timing error in different films, we rank the six films of Series A for each observer according to percentages of discrepantly timed units, from largest to smallest, we find an exact correspondence between size of percentage and film for the *Person* category for all observers. The six films from most to least difficult (difficulty

being measured by amount of timing error) fall in the order I–V–VI–II–IV–III. Observer Y has the same percentage of discrepancy for Films VI and II. The relationship for *Self* is less uniform, but is still quite consistent. Three of the four observers have largest percentages for Film I, next largest for Film V, and all four have the lowest percentage for Film IV. In the *Material* category, there is even less uniformity, but Film III stands either in first or second place for all observers, that is, it has the largest or next largest percentage of error, and Film VI, for three observers, has the smallest, or next smallest, percentage.

To summarize briefly, the size of indices of timing tendency per category was found to vary uniformly for all observers, the largest indices occurring in the *Material* category, the next largest, in *Person*, and the smallest, in *Self*, for Series A. In Series C, the largest percentages were found in *Person*, the next largest, in *Self*, the smallest, in *Material*. Among the factors causing these consistent differences between categories are the amount of behavior recorded per category and the continuity or discontinuity of the particular behavior. Since many of the differences were greater than would be expected, if chance factors alone were operating, we may assume that other factors which as yet we have been unable to distinguish also affect the size of indices per category. A general tendency toward improvement in extent of timing tendency in consecutive observations of the same behavior was noted for all observers. Less improvement was shown in amount of timing error in successive observations of the same films. These percentages tended, on the contrary, to remain relatively constant from the first to the third observation. Percentage indices of timing tendency based on the total record of behavior without regard to interpretation were much smaller than those for the separate categories. The larger bases of the percentages and the greater continuity of the record, when category distinctions were disregarded, presumably operaated to reduce the size of these percentages. Individual differences among films tended to obscure any general trend toward improvement in timing the same types of behavior in different films over a period of weeks. The relative amount of timing error per film in Series A was consistent for all observers in the *Person* category, and fairly consistent in the other two categories.

CHAPTER VIII

THE MEASUREMENT OF INTERPRETIVE DISAGREEMENT

The three categories, *Material*, *Self* and *Person*, which were used in the moving picture observations were defined in such a way as to be mutually exclusive, *Material* taking precedence over both *Self* and *Person*, and *Person* taking precedence over *Self*. The observers recorded the behavior of the film character in terms of these three categories along a time scale divided into five-second intervals. In the subsequent process of transcription to master sheets, the half interval was taken as the limit of accuracy. "Errors" in interpretation may be considered, therefore, as disagreements among observers in the allocation of behavior to different categories in terms of half-interval units.

Crude Measurement of Interpretive Disagreement

The simplest conception of amount of interpretive disagreement for a given observer is the sum of his recorded units which disagree with those of other observers. When this sum is divided by his total recorded units we have an index of interpretive disagreement which is comparable with similar indices for other observers. The disagreements of each observer with each other observer are tabulated in the following six classes:

(1) *Material-Self* (the standard observer records *Material*, the co-observer, *Self*)
(2) *Material-Person* (the standard observer records *Material*, the co-observer, *Person*)
(3) *Self-Material* (the standard observer records *Self*, the co-observer, *Material*)
(4) *Self-Person* (the standard observer records *Self*, the co-observer, *Person*)
(5) *Person-Material* (the standard observer records *Person*, the co-observer, *Material*)
(6) *Person-Self* (the standard observer records *Person*, the co-observer, *Self*)

A unit of disagreement was defined as a unit of record 2½ seconds in length on the time scale throughout which the standard observer recorded one category and the co-observer a different one. Thus, by definition, disagreements were confined to disagreeing records extending over a full 2½ second unit on the scale. A unit of agreement was defined as a record of behavior by the standard observer extending throughout a 2½ second unit on the scale, the agreement being either with the observer himself or with the co-observer. Since, in the transcription of records, any fraction of a unit was counted as a whole unit, an observer's "units of agreement" included all behavior recorded by him.

Each observer's index of disagreement with each other observer for a particular film and observation is obtained for each of the three categories in the following manner. The number of disagreeing units in classes (1) and (2) is expressed as a percentage of the standard observer's total weighted recorded units in the *Material* category. The weighting is in terms of the co-observer's agreement with the standard observer, i. e., all units in which *Material* is recorded by the co-observer synchronously with the standard observer are weighted 2, whereas those units of *Material* recorded by the standard observer only are wieghted 1.[1] The index so derived is called the standard observer's index of disagreement with the co-observer for the *Material* category. Similar indices are derived for *Self* by combining disagreements in classes (3) and (4) and expressing them as a percentage of the standard observer's weighted units of *Self*. For *Person*, the disagreements in classes (5) and (6) are combined and expressed as a percentage of the standard observer's weighted units of *Person*.

The basic data for each of the three observations of the six films of Series A are presented in Table XV (A, B, C). Table XVI shows the indices of disagreement for each observer as a

[1] Units of agreement in accordance with this definition represent the total units in which the standard observer recorded the observed category, weighted by the number of observers recording synchronously. Thus when the index is computed for an observer compared with one other observer, the minimum "agreement" is agreement with himself, i. e., a unit uncorroborated by the other observer; the maximum agreement is a unit corroborated by the other observer. The former unit is weighted 1, the latter 2. Similarly, if the comparison is with two other observers, the minimum unit weight is 1, the maximum 3, and with three other observers, the minimum unit weight is 1, the maximum 4. These latter comparisons are not considered in this volume.

standard compared with each co-observer. This table may be read as follows: Observer W, paired with X, in the first observation of Film I, Series A, had an index of interpretive disagreement of 6.5% for *Material*, i. e., 6.5% of his total weighted units in the *Material* category were synchronous with units of *Person* or *Self* on Observer X's record. In the second observation of the same film and category his index of disagreement, with Observer X as the co-observer, was 5.6%, and in the third observation, 7.2%.

Table XVII presents indices for each standard observer, computed on the basis of his total disagreements with his three co-observers. These indices were computed in a manner exactly comparable with that used in the computation of the single indices.

These basic tables are analyzed in order to throw light on the following points.

(1) What is the degree of disagreement among observers for each category considered separately?

(2) How do observers differ from one another in the degree of interpretive disagreement?

(3) Do the interpretive disagreements of the observers diminish with consecutive observations? If so, what is the character of the improvement? Is it regular, irregular and influenced by chance, or irregular with a definite bias?

(4) How much variation is there in the several pairwise combinations of observers?

(5) How does an observer who deviates widely from other observers affect the indices of interpretive disagreement of the co-observers?

Extent of Interpretive Disagreement per Category of Behavior

The distribution, by category, of the indices of interpretive disagreement for each observer paired with each other observer in each observation of the six films of Series A is shown graphically in Histogram Series IV. It is clear that there is a wide difference in size of index for different categories.

The indices for the *Material* category (Table XVI) show the widest variation, ranging from 0 to 18.9%, with the median index at 5.2%. The indices for the *Self* category cluster closely in the

TABLE XV—*A*

Total Number of Recorded Units of Behavior and Number of Disagreeing Units in the MATERIAL *Category, by Film, Observation and Pairwise Combination of Observers (Series A)*

Observers			Film I				Film II				Film III				Film IV				Film V				Film VI				Films I–VI			
				Units of Disagreement				Units of Disagreement				Units of Disagreement				Units of Disagreement				Units of Disagreement				Units of Disagreement				Units of Disagreement		
Standard Observer	Co-observer	Observation	Total Units	M-S	M-P	M-SP	Total Units	M-S	M-P	M-SP	Total Units	M-S	M-P	M-SP	Total Units	M-S	M-P	M-SP	Total Units	M-S	M-P	M-SP	Total Units	M-S	M-P	M-SP	Total Units	M-S	M-P	M-SP
W	X	1	107	6	1	7	185	29	6	35	60	3	0	3	155	9	0	9	209	12	4	16	245	16	4	20	961	75	15	90
		2	142	4	4	8	214	9	6	15	74	4	0	4	174	9	0	9	218	6	0	6	363	24	3	27	1185	56	13	69
		3	138	8	2	10	216	8	3	11	72	3	0	3	176	3	6	9	251	2	0	2	364	6	2	8	1217	30	13	43
		Total	387	18	7	25	615	46	15	61	206	10	0	10	505	21	6	27	678	20	4	24	972	46	9	55	3363	161	41	202
	Y	1	121	1	2	3	203	17	7	24	63	4	0	4	153	3	5	8	203	12	12	24	226	19	5	24	969	56	31	87
		2	152	0	2	2	231	8	2	10	80	0	0	0	178	9	2	11	205	12	7	19	353	35	7	42	1199	64	20	84
		3	158	1	2	3	230	2	3	5	77	2	0	2	184	2	4	6	238	5	3	8	355	9	3	12	1242	21	15	36
		Total	431	2	6	8	664	27	12	39	220	6	0	6	515	14	11	25	646	29	22	51	934	63	15	78	3410	141	66	207
	Z	1	108	4	6	10	184	15	9	24	62	3	0	3	154	2	0	2	209	9	5	14	248	6	5	11	965	39	25	64
		2	135	3	3	6	219	5	12	17	73	2	1	3	168	8	5	13	217	2	5	7	368	26	4	30	1180	46	30	76
		3	147	7	1	8	222	5	3	8	73	3	0	3	184	0	0	0	238	1	0	1	355	13	2	15	1219	29	6	35
		Total	390	14	10	24	625	25	24	49	208	8	1	9	506	10	5	15	664	12	10	22	971	45	11	56	3364	114	61	175
	XYZ	1	336	11	9	20	572	61	22	83	185	10	0	10	462	14	5	19	621	33	21	54	719	41	14	55	2895	170	71	241
		2	429	7	9	16	664	22	20	42	227	6	1	7	520	26	7	33	640	20	12	32	1084	85	14	99	3654	166	63	229
		3	443	16	5	21	668	15	9	24	222	8	0	8	544	5	10	15	727	8	3	11	1074	28	7	35	3678	80	34	114
		Total	1208	34	23	57	1904	98	51	149	634	24	1	25	1526	45	22	67	1988	61	36	97	2877	154	35	189	10137	416	168	584
X	W	1	113	5	2	7	145	6	1	7	65	4	0	4	163	6	0	6	204	4	2	6	262	16	23	39	952	41	28	69
		2	140	3	1	4	183	5	0	5	70	1	0	1	170	2	0	2	212	2	1	3	341	10	6	16	1116	23	8	31
		3	139	6	4	10	194	2	1	3	73	1	1	2	175	5	1	6	257	5	1	6	354	6	3	9	1192	25	11	36
		Total	392	14	7	21	522	13	2	15	208	6	1	7	508	13	1	14	673	11	4	15	957	32	32	64	3260	89	47	136
	Y	1	125	3	4	7	138	6	6	12	70	3	1	4	172	3	3	6	193	9	10	19	245	26	20	46	943	50	44	94
		2	152	0	1	1	185	3	0	3	71	3	0	3	167	6	0	6	197	8	8	16	328	18	5	23	1100	38	14	52
		3	152	2	3	5	195	0	0	0	78	1	1	2	176	5	1	6	241	11	2	13	328	15	2	17	1170	34	9	43
		Total	429	5	8	13	518	9	6	15	219	7	2	9	515	14	4	18	631	28	20	48	901	59	27	86	3213	122	67	189
	Z	1	115	3	3	6	137	9	1	10	65	7	0	7	166	4	0	4	200	7	2	9	277	9	13	22	960	39	19	58
		2	134	2	2	4	174	5	5	10	67	3	1	4	163	4	6	10	209	0	4	4	337	7	3	10	1084	21	21	42
		3	138	6	5	11	188	4	3	7	75	0	0	0	178	2	0	2	251	7	1	8	338	13	2	15	1168	32	11	43
		Total	387	11	10	21	499	18	9	27	207	10	1	11	507	10	6	16	660	14	7	21	952	29	18	47	3212	92	51	143
	WYZ	1	353	11	9	20	420	21	8	29	200	14	1	15	501	13	3	16	597	20	14	34	784	51	56	107	2855	130	91	221
		2	426	5	4	9	542	13	5	18	208	7	1	8	500	12	6	18	618	10	13	23	1006	35	14	49	3300	82	43	125
		3	429	14	12	26	577	6	4	10	226	2	2	4	529	12	2	14	749	23	4	27	1020	34	7	41	3530	91	31	122
		Total	1208	30	25	55	1539	40	17	57	634	23	4	27	1539	37	11	48	1964	53	31	84	2810	120	77	197	9685	303	165	468

TABLE XV—*A*—*Continued*

Y	W	1	166	22	3	25	194	13	1	14	72	1	0	1	174	15	0	15	192	10	2	12	194	6	5	11	992	67	11	78
		2	179	11	1	12	231	6	0	6	86	3	1	4	189	10	1	11	196	11	0	11	315	9	4	13	1196	50	7	57
		3	177	6	0	6	232	5	2	7	87	3	2	5	190	4	0	4	234	5	0	5	353	8	7	15	1273	31	11	42
		Total	522	39	4	43	657	24	3	27	245	7	3	10	553	29	1	30	622	26	2	28	862	23	16	39	3461	148	29	177
	X	1	164	25	3	28	169	26	4	30	74	2	0	2	185	13	0	13	187	6	1	7	196	12	4	16	975	84	12	96
		2	181	11	2	13	216	11	4	15	81	6	0	6	182	17	0	17	194	9	1	10	312	4	0	4	1166	58	7	65
		3	170	14	0	14	219	16	5	21	87	4	1	5	183	7	5	12	231	6	1	7	343	11	7	18	1233	58	19	77
		Total	515	50	5	55	604	53	13	66	242	12	1	13	550	37	5	42	612	21	3	24	851	27	11	38	3374	200	38	238
	Z	1	173	24	3	27	175	12	6	18	74	3	0	3	190	4	0	4	191	10	1	11	192	8	6	14	995	61	16	77
		2	176	11	4	15	223	5	9	14	81	4	1	5	182	9	4	13	196	8	1	9	314	10	2	12	1172	47	21	68
		3	174	12	3	15	228	13	3	16	88	1	0	1	194	0	0	0	230	3	1	4	344	10	8	18	1258	39	15	54
		Total	523	47	10	57	626	30	18	48	243	8	1	9	566	13	4	17	617	21	3	24	850	28	16	44	3425	147	52	199
	WXZ	1	503	71	9	80	538	51	11	62	220	6	0	6	549	32	0	32	570	26	4	30	582	26	15	41	2962	212	39	251
		2	536	33	7	40	670	22	13	35	248	13	2	15	553	36	5	41	586	28	2	30	941	23	6	29	3534	155	35	190
		3	521	32	3	35	679	34	10	44	262	8	3	11	567	11	5	16	695	14	2	16	1040	29	22	51	3764	128	45	173
		Total	1560	136	19	155	1887	107	34	141	730	27	5	32	1669	79	10	89	1851	68	8	76	2563	78	43	121	10260	495	119	614
Z	W	1	127	17	0	17	158	11	1	12	70	8	0	8	176	14	0	14	223	11	2	13	267	17	10	27	1021	78	13	91
		2	142	9	0	9	216	7	1	8	71	0	0	0	163	5	0	5	234	8	5	13	351	7	2	9	1177	36	8	44
		3	149	4	2	6	219	6	3	9	76	5	1	6	199	8	0	8	253	5	3	8	344	7	0	7	1240	35	9	44
		Total	418	30	2	32	593	24	5	29	217	13	1	14	538	27	0	27	710	24	10	34	962	31	12	43	3438	149	30	179
	X	1	128	16	0	16	151	15	3	18	68	7	0	7	180	11	0	11	219	14	2	16	279	16	2	18	1025	79	7	86
		2	143	6	2	8	202	17	2	19	69	2	0	2	162	10	0	10	232	10	4	14	342	11	3	14	1150	56	11	67
		3	139	4	1	5	207	10	6	16	77	4	0	4	194	9	4	13	260	7	3	10	337	7	0	7	1214	41	14	55
		Total	410	26	3	29	560	42	11	53	214	13	0	13	536	30	4	34	711	31	9	40	958	34	5	39	3389	176	32	208
	Y	1	147	6	0	6	158	8	3	11	73	5	0	5	191	4	5	9	216	17	8	25	243	30	12	42	1028	70	28	98
		2	156	4	2	6	220	8	1	9	73	0	0	0	166	5	0	5	222	15	10	25	335	14	6	20	1172	46	19	65
		3	157	3	1	4	223	5	2	7	85	1	1	2	203	5	3	8	249	11	5	16	335	12	2	14	1252	37	14	51
		Total	460	13	3	16	601	21	6	27	231	6	1	7	560	14	8	22	687	43	23	66	913	56	20	76	3452	153	61	214
	WXY	1	402	39	0	39	467	34	7	41	211	20	0	20	547	29	5	34	658	42	12	54	789	63	24	87	3074	227	48	275
		2	441	19	4	23	638	32	4	36	213	2	0	2	491	20	0	20	688	33	19	52	1028	32	11	43	3499	138	38	176
		3	445	11	4	15	649	21	11	32	238	10	2	12	596	22	7	29	762	23	11	34	1016	26	2	28	3706	113	37	150
		Total	1288	69	8	77	1754	87	22	109	662	32	2	34	1634	71	12	83	2108	98	42	140	2833	121	37	158	10279	478	123	601
WXYZ		1	1594	132	27	159	1997	167	48	215	816	50	1	51	2059	88	13	101	2446	121	51	172	2874	181	109	290	11786	739	249	988
		2	1832	64	24	88	2514	89	42	131	896	28	4	32	2064	94	18	112	2532	91	46	137	4059	175	45	220	13897	541	179	720
		3	1838	73	24	97	2573	76	34	110	948	28	7	35	2236	50	24	74	2933	68	20	88	4150	117	38	155	14678	412	147	559
		Total	5264	269	75	344	7084	332	124	456	2660	106	12	118	6359	232	55	287	7911	280	117	397	11083	473	192	665	40361	1692	575	2267

Notes: (1) Total units = "units of agreement." (2) All units weighted by the number of persons recording.

TABLE XV—*B*

Total Number of Recorded Units of Behavior and Number of Disagreeing Units in the SELF *Category, by Film, Observation and Pairwise Combination of Observers (Series A)*

Observers			Film I				Film II				Film III				Film IV				Film V				Film VI				Films I–VI			
				Units of Disagreement				Units of Disagreement				Units of Disagreement				Units of Disagreement				Units of Disagreement				Units of Disagreement				Units of Disagreement		
Standard Observer	Co-observers	Observation	Total Units	S-M	S-P	S-MP	Total Units	S-M	S-P	S-MP	Total Units	S-M	S-P	S-MP	Total Units	S-M	S-P	S-MP	Total Units	S-M	S-P	S-MP	Total Units	S-M	S-P	S-MP	Total Units	S-M	S-P	S-MP
W	X	1	1054	5	5	10	1684	6	12	18	1165	4	2	6	1990	6	11	17	1441	4	1	5	1681	16	20	36	9015	41	51	92
		2	1002	3	6	9	1699	5	16	21	1198	1	6	7	2004	2	8	10	1442	2	6	8	1661	10	17	27	9006	23	59	82
		3	1036	6	3	9	1701	2	6	8	1210	1	2	3	2018	5	9	14	1494	5	6	11	1694	6	17	23	9153	25	43	68
		Total	3092	14	14	28	5084	13	34	47	3573	6	10	16	6012	13	28	41	4377	11	13	24	5036	32	54	86	27174	89	153	242
	Y	1	1009	22	22	44	1659	13	23	36	1169	1	2	3	1969	15	10	25	1420	10	10	20	1701	6	11	17	8927	67	78	145
		2	971	11	7	18	1710	6	17	23	1203	3	6	9	1972	10	10	20	1451	11	9	20	1653	9	25	34	8960	50	74	124
		3	1017	6	13	19	1680	5	18	23	1208	3	5	8	2025	4	10	14	1491	5	8	13	1711	8	12	20	9132	31	66	97
		Total	2997	39	42	81	5049	24	58	82	3580	7	13	20	5966	29	30	59	4362	26	27	53	5065	23	48	71	27019	148	218	366
	Z	1	1000	17	23	40	1619	11	34	45	1156	8	4	12	1924	14	18	32	1371	11	8	19	1620	17	38	55	8690	78	125	203
		2	960	9	9	18	1667	7	31	38	1180	0	10	10	1962	5	12	17	1397	8	13	21	1609	7	22	29	8775	36	97	133
		3	1029	4	6	10	1673	6	16	22	1188	5	3	8	1980	8	10	18	1451	5	7	12	1655	7	22	29	8976	35	64	99
		Total	2989	30	38	68	4959	24	81	105	3524	13	17	30	5866	27	40	67	4219	24	28	52	4884	31	82	113	26441	149	286	435
	XYZ	1	3063	44	50	94	4962	30	69	99	3490	13	8	21	5883	35	39	74	4232	25	19	44	5002	39	69	108	26632	186	254	440
		2	2933	23	22	45	5076	18	64	82	3581	4	22	26	5938	17	30	47	4290	21	28	49	4923	26	64	90	26741	109	230	339
		3	3082	16	22	38	5054	13	40	53	3606	9	10	19	6023	17	29	46	4436	15	21	36	5060	21	51	72	27261	91	173	264
		Total	9078	83	94	177	15092	61	173	234	10677	26	40	66	17844	69	98	167	12958	61	68	129	14985	86	184	270	80634	386	657	1043
X	W	1	1042	6	4	10	1730	29	26	55	1189	3	2	5	1997	9	3	12	1443	12	4	16	1672	16	11	27	9073	75	50	125
		2	1017	4	9	13	1681	9	14	23	1180	4	5	9	2007	9	8	17	1439	6	4	10	1674	24	12	36	8998	56	52	108
		3	1064	8	12	20	1700	8	14	22	1222	3	2	5	1992	3	2	5	1492	2	1	3	1687	6	8	14	9157	30	39	69
		Total	3132	18	25	43	5111	46	54	100	3591	10	9	19	5996	21	13	34	4374	20	9	29	5033	46	31	77	27228	161	141	302
	Y	1	1000	25	21	46	1721	26	28	54	1211	2	3	5	1988	13	10	23	1423	6	11	17	1677	12	16	28	9020	84	89	173
		2	996	11	12	23	1675	11	21	32	1188	6	0	6	1971	17	14	31	1442	9	9	18	1685	4	23	27	8957	58	79	137
		3	1056	14	13	27	1681	16	22	38	1228	4	4	8	1990	7	6	13	1491	6	0	6	1675	11	16	27	9121	58	61	119
		Total	3052	50	46	96	5077	53	71	124	3627	12	7	19	5949	37	30	67	4356	21	20	41	5037	27	55	82	27098	200	229	429
	Z	1	989	16	22	38	1662	15	46	61	1194	7	3	10	1935	11	19	30	1374	14	12	26	1602	16	22	38	8756	79	124	203
		2	985	6	16	22	1635	17	35	52	1162	2	3	5	1953	10	12	22	1384	10	14	24	1637	11	21	32	8756	56	101	157
		3	1058	4	10	14	1668	10	22	32	1201	4	2	6	1936	9	4	13	1434	7	6	13	1639	7	25	32	8936	41	69	110
		Total	3032	26	48	74	4965	42	103	145	3557	13	8	21	5824	30	35	65	4192	31	32	63	4878	34	68	102	26448	176	294	470
	WYZ	1	3031	47	47	94	5113	70	100	170	3594	12	8	20	5920	33	32	65	4240	32	27	59	4951	44	49	93	26849	238	263	501
		2	2998	21	37	58	4991	37	70	107	3530	12	8	20	5931	36	34	70	4265	25	27	52	4996	39	56	95	26711	170	232	402
		3	3178	26	35	61	5049	34	58	92	3651	11	8	19	5918	19	12	31	4417	15	7	22	5001	24	49	73	27214	129	169	298
		Total	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

TABLE XV—B—*Continued*

Y	W	1	941	1	0	1	1630	17	6	23	1196	4	4	8	1944	3	9	12	1409	12	3	15	1722	19	22	41	8842	56	44	100
		2	942	0	3	3	1713	8	17	25	1191	0	9	9	1944	9	6	15	1468	12	2	14	1676	35	18	53	8934	64	55	119
		3	1000	1	4	5	1640	2	10	12	1216	2	4	6	2006	2	2	4	1497	5	4	9	1715	9	12	21	9074	21	36	57
		Total	2883	2	7	9	4983	27	33	60	3603	6	17	23	5894	14	17	31	4374	29	9	38	5113	63	52	115	26850	141	135	276
	X	1	944	3	3	6	1646	6	5	11	1214	3	3	6	1956	3	10	13	1410	9	6	15	1707	26	28	54	8877	50	55	105
		2	952	0	5	5	1696	3	21	24	1194	3	4	7	1940	6	7	13	1462	8	5	13	1695	18	22	40	8939	38	64	102
		3	1011	2	3	5	1642	0	2	2	1224	1	3	4	1997	5	4	9	1499	11	2	13	1686	15	22	37	9059	34	36	70
		Total	2907	5	11	16	4984	9	28	37	3632	7	10	17	5893	14	21	35	4371	28	13	41	5088	59	72	131	26875	122	155	277
	Z	1	896	6	19	25	1572	8	22	30	1204	5	1	6	1895	4	11	15	1356	17	9	26	1646	30	40	70	8569	70	102	172
		2	912	4	10	14	1670	8	30	38	1176	0	5	5	1898	5	10	15	1427	15	7	22	1656	14	25	39	8739	46	87	133
		3	996	3	3	6	1611	5	9	14	1204	1	4	5	1954	5	2	7	1456	11	4	15	1657	12	25	37	8878	37	47	84
		Total	2804	13	32	45	4853	21	61	82	3584	6	10	16	5747	14	23	37	4239	43	20	63	4959	56	90	146	26186	153	236	389
	WXZ	1	2781	10	22	32	4848	31	33	64	3614	12	8	20	5795	10	30	40	4175	38	18	56	5075	75	90	165	26288	176	201	377
		2	2806	4	18	22	5079	19	68	87	3561	3	18	21	5782	20	23	43	4357	35	14	49	5027	67	65	132	26612	148	206	354
		3	3007	6	10	16	4893	7	21	28	3644	4	11	15	5957	12	8	20	4452	27	10	37	5058	36	59	95	27011	92	119	211
		Total	8594	20	50	70	14820	57	122	179	10819	19	37	56	17534	42	61	103	12984	100	42	142	15160	178	214	392	79911	416	526	942
Z	W	1	937	4	3	7	1555	15	14	29	1179	3	5	8	1871	2	14	16	1327	9	2	11	1534	6	17	23	8403	39	55	94
		2	944	3	2	5	1610	5	5	10	1149	2	4	6	1917	8	8	16	1349	2	5	7	1572	26	9	35	8541	46	33	79
		3	1037	7	10	17	1662	5	13	18	1168	3	2	5	1907	0	7	7	1412	1	6	7	1611	13	10	23	8797	29	48	77
		Total	2918	14	15	29	4827	25	32	57	3496	8	11	19	5695	10	29	39	4088	12	13	25	4717	45	36	81	25741	114	136	250
	X	1	938	3	5	8	1552	9	8	17	1193	7	4	11	1875	3	15	18	1328	7	3	10	1525	9	12	21	8411	38	47	85
		2	954	2	2	4	1596	5	9	14	1149	3	0	3	1905	4	12	16	1339	0	5	5	1587	7	14	21	8530	21	42	63
		3	1038	6	7	13	1658	4	7	11	1169	0	1	1	1889	2	10	12	1397	7	9	16	1602	13	14	27	8753	32	48	80
		Total	2930	11	14	25	4806	18	24	42	3511	10	5	15	5669	9	37	46	4064	14	17	31	4714	29	40	69	25694	91	137	228
	Y	1	901	24	17	41	1537	12	18	30	1200	3	3	6	1867	4	10	14	1323	10	4	14	1539	8	20	28	8367	61	72	133
		2	925	11	10	21	1610	5	5	10	1157	4	0	4	1881	9	14	23	1362	8	3	11	1596	10	19	29	8531	47	51	98
		3	1021	12	8	20	1640	13	14	27	1176	1	0	1	1900	0	11	11	1411	3	8	11	1609	10	10	20	8757	39	51	90
		Total	2847	47	35	82	4787	30	37	67	3533	8	3	11	5648	13	35	48	4096	21	15	36	4744	28	49	77	25655	147	174	321
	WXY	1	2776	31	25	56	4644	36	40	76	3572	13	12	25	5613	9	39	48	3978	26	9	35	4598	23	49	72	25181	138	174	312
		2	2823	16	14	30	4816	15	19	34	3455	9	4	13	5703	21	34	55	4050	10	13	23	4755	43	42	85	25602	114	126	240
		3	3096	25	25	50	4960	22	34	56	3513	4	3	7	5696	2	28	30	4220	11	23	34	4822	36	34	70	26307	100	147	247
		Total	8695	72	64	136	14420	73	93	166	10540	26	19	45	17012	32	101	133	12248	47	45	92	14175	102	125	227	77090	352	447	799
WXYZ		1	11651	132	144	276	19567	167	242	409	14270	50	36	86	23211	87	140	227	16625	121	73	194	19626	181	257	438	104950	738	892	1630
		2	11560	64	91	155	19962	89	221	310	14127	28	52	80	23354	94	121	215	16962	91	82	173	19701	175	227	402	105666	541	794	1335
		3	12363	73	92	165	19956	76	153	229	14414	28	32	60	23594	50	77	127	17525	68	61	129	19941	117	193	310	107793	412	608	1020
		Total	35574	269	327	596	59485	332	616	948	42811	106	120	226	70159	231	338	569	51112	280	216	496	59268	473	677	1150	318409	1691	2294	3985

Notes: (1) Total units = "units of agreement." (2) All units weighted by number of persons recording.

TABLE XV—C

Total Number of Recorded Units of Behavior and Number of Disagreeing Units in the PERSON *Category, by Film, Observation and Pairwise Combination of Observers (Series A)*

OBSERVERS			FILM I				FILM II				FILM III				FILM IV				FILM V				FILM VI				FILMS I–VI			
Standard Observer	Co-observer	Observation	Total Units	Units of Disagreement P-M	P-S	P-MS	Total Units	Units of Disagreement P-M	P-S	P-MS	Total Units	Units of Disagreement P-M	P-S	P-MS	Total Units	Units of Disagreement P-M	P-S	P-MS	Total Units	Units of Disagreement P-M	P-S	P-MS	Total Units	Units of Disagreement P-M	P-S	P-MS	Total Units	Units of Disagreement P-M	P-S	P-MS
W	X	1	223	2	4	6	532	1	26	27	173	0	2	2	478	0	3	3	163	2	4	6	392	23	11	34	1961	28	50	78
		2	256	1	9	10	517	0	14	14	186	0	5	5	466	0	8	8	162	1	4	5	409	6	12	18	1996	8	52	60
		3	256	4	12	16	584	1	14	15	174	1	2	3	465	1	2	3	171	1	1	2	410	3	8	11	2060	11	39	50
		Total	735	7	25	32	1633	2	54	56	533	1	9	10	1409	1	13	14	496	4	9	13	1211	32	31	63	6017	47	141	188
	Y	1	235	3	0	3	563	1	6	7	171	0	4	4	479	0	9	9	169	2	3	5	383	5	22	27	2000	11	44	55
		2	277	1	3	4	545	0	17	17	184	1	9	10	469	1	6	7	173	0	2	2	404	4	18	22	2052	7	55	62
		3	278	0	4	4	616	2	10	12	177	2	4	6	465	0	2	2	172	0	4	4	409	7	12	19	2117	11	36	47
		Total	790	4	7	11	1724	3	33	36	532	3	17	20	1413	1	17	18	514	2	9	11	1196	16	52	68	6169	29	135	164
	Z	1	226	0	3	3	534	1	14	15	172	0	5	5	466	0	14	14	166	2	2	4	376	10	17	27	1940	13	55	68
		2	275	0	2	2	556	1	5	6	187	0	4	4	458	0	8	8	162	5	5	10	419	2	9	11	2057	8	33	41
		3	272	2	10	12	587	3	13	16	177	1	2	3	453	0	7	7	165	3	6	9	417	0	10	10	2071	9	48	57
		Total	773	2	15	17	1677	5	32	37	536	1	11	12	1377	0	29	29	493	10	13	23	1212	12	36	48	6068	30	136	166
	XYZ	1	684	5	7	12	1629	3	46	49	516	0	11	11	1423	0	26	26	498	6	9	15	1151	38	50	88	5901	52	149	201
		2	808	2	14	16	1618	1	36	37	557	1	18	19	1393	1	22	23	497	6	11	17	1232	12	39	51	6105	23	140	163
		3	806	6	26	32	1787	6	37	43	528	4	8	12	1383	1	11	12	508	4	11	15	1236	10	30	40	6248	31	123	154
		Total	2298	13	47	60	5034	10	119	129	1601	5	37	42	4199	2	59	61	1503	16	31	47	3619	60	119	179	18254	106	412	518
	W	1	229	1	5	6	536	6	12	18	180	0	2	2	479	0	11	11	156	4	1	5	400	4	20	24	1980	15	51	66
		2	240	4	6	10	517	6	16	22	190	0	6	6	463	0	8	8	157	0	6	6	410	3	17	20	1977	13	59	72
		3	240	2	3	5	565	3	6	9	174	0	2	2	484	6	9	15	175	0	6	6	414	2	17	19	2052	13	43	56
		Total	709	7	14	21	1618	15	34	49	544	0	10	10	1426	6	28	34	488	4	13	17	1224	9	54	63	6009	41	153	194
	Y	1	238	3	3	6	563	4	5	9	187	0	3	3	485	0	10	10	155	1	6	7	385	4	28	32	2013	12	55	67
		2	249	2	5	7	538	4	21	25	197	0	4	4	461	0	7	7	163	1	5	6	416	0	22	22	2024	7	64	71
		3	255	0	3	3	567	5	2	7	197	1	3	4	496	5	4	9	180	1	2	3	408	7	22	29	2103	19	36	55
		Total	742	5	11	16	1668	13	28	41	581	1	10	11	1442	5	21	26	498	3	13	16	1209	11	72	83	6140	38	155	193
	Z	1	234	0	5	5	538	3	8	11	186	0	4	4	465	0	15	15	155	2	3	5	399	2	12	14	1977	7	47	54
		2	239	2	2	4	548	2	9	11	200	0	0	0	452	0	12	12	155	4	5	9	419	3	14	17	2013	11	42	53
		3	241	1	7	8	553	6	7	13	175	0	1	1	479	4	10	14	172	3	9	12	423	0	14	14	2043	14	48	62
		Total	714	3	14	17	1639	11	24	35	561	0	5	5	1396	4	37	41	482	9	17	26	1241	5	40	45	6033	32	137	169
	WYZ	1	701	4	13	17	1637	13	25	38	553	0	9	9	1429	0	36	36	466	7	10	17	1184	10	60	70	5970	34	153	187
		2	728	8	13	21	1603	12	46	58	587	0	10	10	1376	0	27	27	475	5	16	21	1245	6	53	59	6014	31	165	196
		3	736	3	13	16	1685	14	15	29	546	1	6	7	1459	15	23	38	527	4	17	21	1245	9	53	62	6198	46	127	173
		Total	2165	15	39	54	4925	39	86	125	1686	1	25	26	4264	15	86	101	1468	16	43	59	3674	25	166	191	18182	111	445	556

TABLE XV—*C*—*Continued*

Y	W	1	270	2	22	24	617	7	23	30	179	0	2	2	499	5	10	15	177	12	10	22	387	5	11	16	2129	31	78	109
		2	288	2	7	9	564	2	17	19	186	0	6	6	493	2	10	12	193	7	9	16	417	7	25	32	2141	20	74	94
		3	291	2	13	15	622	3	18	21	185	0	5	5	481	4	10	14	182	3	8	11	411	3	12	15	2172	15	66	81
		Total	849	6	42	48	1803	12	58	70	550	0	13	13	1473	11	30	41	552	22	27	49	1215	15	48	63	6442	66	218	284
	X	1	267	4	21	25	613	6	28	34	188	1	3	4	504	3	10	13	170	10	11	21	381	20	16	36	2123	44	89	133
		2	276	1	12	13	557	0	21	21	195	0	0	0	488	0	14	14	188	8	9	17	428	5	23	28	2132	14	79	93
		3	284	3	13	16	602	0	22	22	184	1	4	5	493	1	6	7	186	2	0	2	406	2	16	18	2155	9	61	70
		Total	827	8	46	54	1772	6	71	77	567	2	7	9	1485	4	30	34	544	20	20	40	1215	27	55	82	6410	67	229	296
	Z	1	276	0	17	17	616	3	18	21	190	0	3	3	496	5	10	15	177	8	4	12	382	12	20	32	2137	28	72	100
		2	283	2	10	12	587	1	5	6	198	0	0	0	488	0	14	14	193	10	3	13	435	6	19	25	2184	19	51	70
		3	295	1	8	9	615	2	14	16	185	1	0	1	481	3	11	14	184	5	8	13	416	2	10	12	2176	14	51	65
		Total	854	3	35	38	1818	6	37	43	573	1	3	4	1465	8	35	43	554	23	15	38	1233	20	49	69	6497	61	174	235
	WXZ	1	813	6	60	66	1846	16	69	85	557	1	8	9	1499	13	30	43	524	30	25	55	1150	37	47	84	6389	103	239	342
		2	847	5	29	34	1708	3	43	46	579	0	6	6	1469	2	38	40	574	25	21	46	1280	18	67	85	6457	53	204	257
		3	870	6	34	40	1839	5	54	59	554	2	9	11	1455	8	27	35	552	10	16	26	1233	7	38	45	6503	38	178	216
		Total	2530	17	123	140	5393	24	166	190	1690	3	23	26	4423	23	95	118	1650	65	62	127	3663	62	152	214	19349	194	621	815
Z	W	1	262	6	23	29	595	9	34	43	180	0	4	4	472	0	18	18	172	5	8	13	392	5	38	43	2073	25	125	150
		2	292	3	9	12	618	12	31	43	198	1	10	11	470	5	12	17	169	5	13	18	435	4	22	26	2182	30	97	127
		3	276	1	6	7	591	3	16	19	181	0	3	3	457	0	10	10	166	0	7	7	443	2	22	24	2114	6	64	70
		Total	830	10	38	48	1804	24	81	105	559	1	17	18	1399	5	40	45	507	10	28	38	1270	11	82	93	6369	61	286	347
	X	1	264	3	22	25	595	1	46	47	187	0	3	3	470	0	19	19	168	2	12	14	407	13	22	35	2091	19	124	143
		2	272	2	16	18	610	5	35	40	207	1	3	4	467	6	12	18	167	4	14	18	434	3	21	24	2157	21	101	122
		3	261	5	10	15	576	3	22	25	179	0	2	2	464	0	4	4	169	1	6	7	445	2	25	27	2094	11	69	80
		Total	797	10	48	58	1781	9	103	112	573	1	8	9	1401	6	35	41	504	7	32	39	1286	18	68	86	6342	51	294	345
	Y	1	277	3	19	22	623	6	22	28	190	0	1	1	482	0	11	11	175	1	9	10	426	6	40	46	2173	16	102	118
		2	289	4	10	14	630	9	30	39	207	1	5	6	476	4	10	14	180	1	7	8	438	2	25	27	2220	21	87	108
		3	286	3	3	6	603	3	9	12	181	0	4	4	469	0	2	2	175	1	4	5	440	8	25	33	2154	15	47	62
		Total	852	10	32	42	1856	18	61	79	578	1	10	11	1427	4	23	27	530	3	20	23	1304	16	90	106	6547	52	236	288
	WXY	1	803	12	64	76	1813	16	102	118	557	0	8	8	1424	0	48	48	515	8	29	37	1225	24	100	124	6337	60	351	411
		2	853	9	35	44	1858	26	96	122	612	3	18	21	1413	15	34	49	516	10	34	44	1307	9	68	77	6559	72	285	357
		3	823	9	19	28	1770	9	47	56	541	0	9	9	1390	0	16	16	510	2	17	19	1328	12	72	84	6362	32	180	212
		Total	2479	30	118	148	5441	51	245	296	1710	3	35	38	4227	15	98	113	1541	20	80	100	3860	45	240	285	19258	164	816	980
WXYZ		1	3001	27	144	171	6925	48	242	290	2183	1	36	37	5775	13	140	153	2003	51	73	124	4710	109	257	366	24597	249	892	1141
		2	3236	24	91	115	6787	42	221	263	2335	4	52	56	5651	18	121	139	2062	46	82	128	5064	45	227	272	25135	179	794	973
		3	3235	24	92	116	7081	34	153	187	2169	7	32	39	5687	24	77	101	2097	20	61	81	5042	38	193	231	25311	147	608	755
		Total	9472	75	327	402	20793	124	616	740	6687	12	120	132	17113	55	338	393	6162	117	216	333	14816	192	677	869	75043	575	2294	2869

Notes: (1) Total units = "units of agreement." (2) All units weighted by number of persons recording.

TABLE XVI

Percentage Indices of Interpretive Disagreement per Observer Expressing the Ratio Between the Number of Disagreeing Units and the Total Number of Units Recorded, by Film, Observation, Category of Behavior and Pairwise Combination of Observers (Series A) *

Standard Observer	Co-observers	Observation	Material Films I	II	III	IV	V	VI	I–VI	Self Films I	II	III	IV	V	VI	I–VI	Person Films I	II	III	IV	V	VI	I–VI
W	X	1	6.5	18.9	5.0	5.8	7.6	8.2	9.4	0.9	1.1	0.5	0.8	0.3	2.1	1.0	2.7	5.1	1.2	0.6	3.7	8.7	4.0
		2	5.6	7.0	5.4	5.2	2.8	7.4	5.8	0.9	1.2	0.6	0.5	0.6	1.6	0.9	3.9	2.7	2.7	1.7	3.1	4.4	3.0
		3	7.2	5.1	4.2	5.1	0.8	2.2	3.5	0.9	0.5	0.2	0.7	0.7	1.4	0.7	6.2	2.6	1.7	0.6	1.2	2.7	2.4
		Av.	6.5	9.9	4.9	5.3	3.5	5.7	6.0	0.9	0.9	0.4	0.7	0.5	1.7	0.8	4.4	3.4	1.9	1.0	2.6	5.2	3.1
	Y	1	2.5	11.8	6.3	5.2	11.8	10.6	9.0	4.4	2.2	0.3	1.3	1.4	1.0	1.6	1.3	1.2	2.3	1.9	3.0	7.0	2.8
		2	1.3	4.3	0	6.2	9.3	11.9	7.0	1.9	1.3	0.7	1.0	1.4	2.1	1.4	1.4	3.1	5.4	1.5	1.2	5.4	3.0
		3	1.9	2.2	2.6	3.3	3.4	3.4	2.9	1.9	1.4	0.7	0.7	0.9	1.2	1.1	1.4	1.9	3.4	0.4	2.3	4.6	2.2
		Av.	1.8	5.9	2.7	4.9	7.9	8.4	6.1	2.7	1.6	0.6	1.0	1.2	1.4	1.4	1.4	2.1	3.8	1.3	2.1	5.7	2.7
	Z	1	9.3	13.0	4.8	1.3	6.7	4.4	6.6	3.1	2.8	1.0	1.7	1.4	3.4	2.3	1.3	2.8	2.9	3.0	2.4	7.2	3.5
		2	4.4	7.8	4.1	7.7	3.2	8.2	6.4	1.9	2.3	0.8	0.9	1.5	1.8	1.5	0.7	1.1	2.1	1.7	6.2	2.6	2.0
		3	5.4	3.6	4.1	0	0.4	4.2	2.9	1.0	1.3	0.7	0.9	0.8	1.8	1.1	4.0	2.7	1.7	1.5	5.5	2.4	2.8
		Av.	6.2	7.8	4.3	3.0	3.3	5.8	5.2	2.3	2.1	0.9	1.1	1.2	2.3	1.6	2.2	2.2	2.2	2.1	4.7	4.0	2.7
X	W	1	6.2	4.8	6.2	3.7	2.9	14.9	7.2	1.0	3.2	0.4	0.6	1.1	1.6	1.4	2.6	3.4	1.1	2.3	3.2	6.0	3.3
		2	2.9	2.7	1.4	1.2	1.4	4.7	2.8	1.3	1.4	0.8	0.8	0.7	2.1	1.2	4.2	4.7	3.2	1.7	3.8	4.9	3.6
		3	7.2	1.5	2.7	3.4	2.3	2.5	3.0	1.9	1.3	0.4	0.3	0.2	0.8	0.8	2.1	1.6	1.1	3.1	3.4	4.6	2.7
		Av.	5.3	2.9	3.4	2.8	2.2	6.7	4.2	1.4	2.0	0.5	0.6	0.7	1.5	1.1	3.0	3.0	1.8	2.4	3.5	5.1	3.2
	Y	1	5.6	8.7	5.7	3.5	9.8	18.8	10.0	4.6	3.1	0.4	1.2	1.2	1.7	1.9	2.5	1.6	1.6	2.1	4.5	8.3	3.3
		2	0.7	1.6	4.2	3.6	8.1	7.0	4.7	2.3	1.9	0.5	1.6	1.2	1.6	1.5	2.8	4.6	2.0	1.5	3.7	5.3	3.5
		3	3.3	0	2.6	3.4	5.4	5.2	3.7	2.6	2.3	0.7	0.7	0.4	1.6	1.3	1.2	1.2	2.0	1.8	1.7	7.1	2.6
		Av.	3.0	2.9	4.1	3.5	7.6	9.5	5.9	3.1	2.4	0.5	1.1	0.9	1.6	1.6	2.1	2.5	1.9	1.8	3.2	6.9	3.1
	Z	1	5.2	7.3	10.8	2.4	4.5	7.9	6.0	3.8	3.7	0.8	1.6	1.9	2.4	2.3	2.1	2.0	2.2	3.2	3.2	3.5	2.7
		2	3.0	5.7	6.0	6.1	1.9	3.0	3.9	2.2	3.2	0.4	1.1	1.7	2.0	1.8	1.7	2.0	0	2.7	5.8	4.1	2.6
		3	7.7	3.7	0	1.1	3.2	4.4	3.7	1.3	1.9	0.5	0.5	0.9	2.0	1.2	3.3	2.4	0.6	2.9	7.0	3.3	3.0
		Av.	5.4	5.4	5.3	3.1	3.2	4.9	4.5	2.4	2.9	0.6	1.1	1.5	2.1	1.8	2.4	2.1	0.9	2.9	5.4	3.6	2.8

* All units weighted by number of persons recording.

TABLE XVI—*Continued*

Percentage Indices of Interpretive Disagreement per Observer Expressing the Ratio Between the Number of Disagreeing Units and the Total Number of Units Recorded, by Film, Observation, Category of Behavior and Pairwise Combination of Observers (Series A) *

Standard Observer	Co-observers	Observation	Material							Self							Person						
			Films							Films							Films						
			I	II	III	IV	V	VI	I–VI	I	II	III	IV	V	VI	I–VI	I	II	III	IV	V	VI	I–VI
Y	W	1	15.1	7.2	1.4	8.6	6.3	5.7	7.9	0.1	1.4	0.7	0.6	1.1	2.4	1.1	8.9	4.9	1.1	3.0	12.4	4.1	5.1
		2	6.7	2.6	4.7	5.8	5.6	4.1	4.8	0.3	1.5	0.8	0.8	1.0	3.2	1.3	3.1	3.4	3.2	2.4	8.3	7.7	4.4
		3	3.4	3.0	5.8	2.1	2.1	4.2	3.3	0.5	0.7	0.5	0.2	0.6	1.2	0.6	5.2	3.4	2.7	3.1	6.0	3.6	3.7
		Av.	8.2	4.1	4.1	5.4	4.5	4.5	5.1	0.3	1.2	0.6	0.5	0.9	2.2	1.0	5.7	3.9	2.6	2.8	8.9	5.2	4.4
	X	1	17.1	17.8	2.7	7.0	3.7	8.2	9.8	0.6	0.7	0.5	0.7	1.1	3.2	1.2	9.4	5.5	2.1	2.6	12.4	9.4	6.3
		2	7.2	6.9	7.4	9.3	5.2	1.3	5.6	0.5	1.4	0.6	0.7	0.9	2.4	1.1	4.7	3.8	0	2.9	9.0	6.5	4.4
		3	8.2	9.6	5.8	6.6	3.0	5.2	6.2	0.5	0.1	0.3	0.5	0.9	2.2	0.8	5.6	3.7	2.7	1.4	1.1	4.4	3.2
		Av.	10.7	10.9	5.4	7.6	3.9	4.5	7.1	0.6	0.7	0.5	0.6	0.9	2.6	1.0	6.5	4.3	1.6	2.3	7.4	6.7	4.6
	Z	1	15.6	10.3	4.0	2.1	5.7	7.3	7.7	2.8	1.9	0.5	0.8	1.9	4.3	2.0	6.2	3.4	1.6	3.0	6.8	8.4	4.7
		2	8.5	6.3	6.2	7.1	4.6	3.8	5.8	1.5	2.3	0.4	0.8	1.5	2.4	1.5	4.2	1.0	0	2.9	6.7	5.7	3.2
		3	8.6	7.0	1.1	0	1.7	5.2	4.3	0.6	0.9	0.4	0.4	1.0	2.2	0.9	3.1	2.6	0.5	2.9	7.1	2.9	3.0
		Av.	10.9	7.7	3.7	3.0	3.9	5.2	5.8	1.6	1.7	0.4	0.6	1.5	2.9	1.5	4.4	2.4	0.7	2.9	6.8	5.6	3.6
Z	W	1	13.4	7.6	11.4	8.0	5.8	10.1	8.9	0.7	1.9	0.7	0.9	0.8	1.5	1.1	11.1	7.2	2.2	3.8	7.6	11.0	7.2
		2	6.3	3.7	0	3.1	5.6	2.6	3.7	0.5	0.6	0.5	0.8	0.5	2.2	0.9	4.1	7.0	5.6	3.6	10.7	6.0	5.8
		3	4.0	4.1	7.9	4.0	3.2	2.0	3.5	1.6	1.1	0.4	0.4	0.5	1.4	0.9	2.5	3.2	1.7	2.2	4.2	5.4	3.3
		Av.	7.7	4.9	6.4	5.0	4.8	4.5	5.2	1.0	1.2	0.5	0.7	0.6	1.7	1.0	5.8	5.8	3.2	3.2	7.5	7.3	5.4
	X	1	12.5	11.9	10.3	6.1	7.3	6.5	8.4	0.8	1.1	0.9	1.0	0.8	1.4	1.0	9.5	7.9	1.6	4.0	8.3	8.6	6.8
		2	5.6	9.4	2.9	6.2	6.0	4.1	5.8	0.4	0.9	0.3	0.8	0.4	1.3	0.7	6.6	6.6	1.9	3.9	10.8	5.5	5.7
		3	3.6	7.7	5.2	6.7	3.8	2.1	4.5	1.3	0.7	0.1	0.6	1.1	1.7	0.9	5.7	4.3	1.1	0.9	4.1	6.1	3.8
		Av.	7.1	9.5	6.1	6.3	5.6	4.1	6.1	0.9	0.9	0.4	0.8	0.8	1.5	0.9	7.3	6.3	1.6	2.9	7.7	6.7	5.4
	Y	1	4.1	7.0	6.8	4.7	11.6	17.3	9.5	4.6	2.0	0.5	0.7	1.1	1.8	1.6	7.9	4.5	0.5	2.3	5.7	10.8	5.4
		2	3.8	4.1	0	3.0	11.3	6.0	5.5	2.3	0.6	0.3	1.2	0.8	1.8	1.2	4.8	6.2	2.9	2.9	4.4	6.2	4.9
		3	2.5	3.1	2.4	3.9	6.4	4.2	4.1	2.0	1.6	0.1	0.6	0.8	1.2	1.0	2.1	2.0	2.2	0.4	2.9	7.5	2.9
		Av.	3.5	4.5	3.0	3.9	9.6	8.3	6.2	2.9	1.4	0.3	0.8	0.9	1.6	1.3	4.9	4.3	1.9	1.9	4.3	8.1	4.4

* All units weighted by number of persons recording.

TABLE XVII

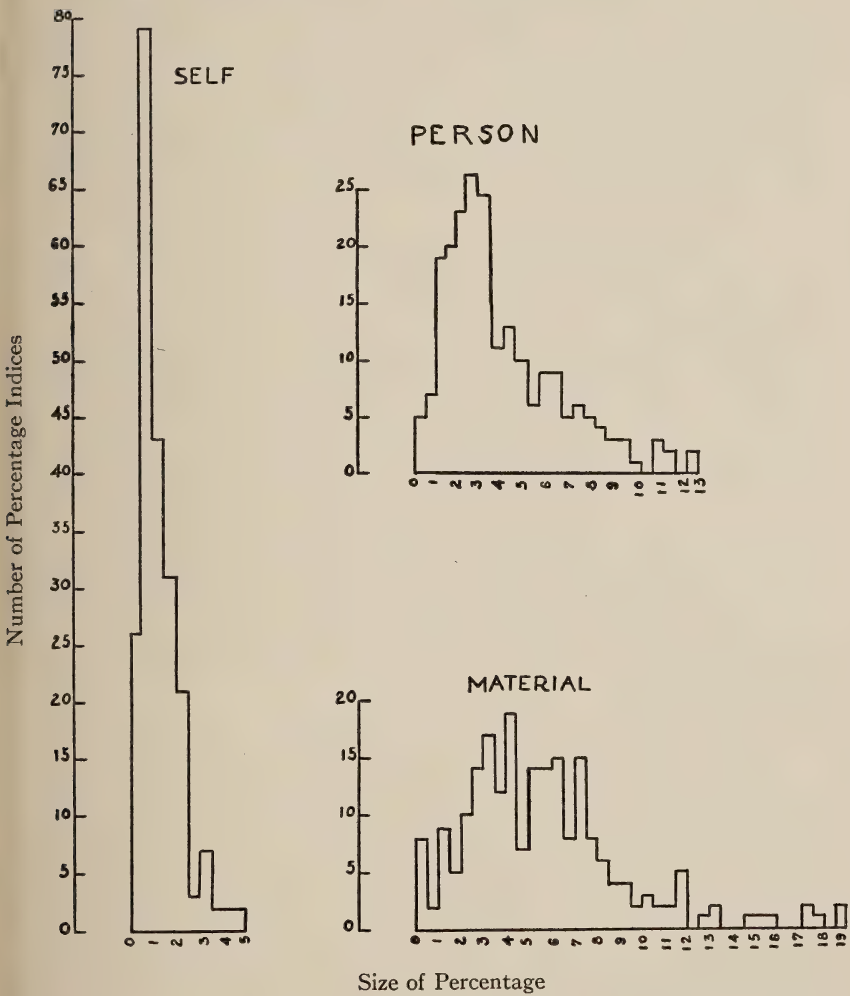
Percentage Indices of Interpretive Disagreement Expressing the Ratio between the Sum of All Disagreeing Units and the Sum of All Units Recorded, for Each Observer in Three Pairings with Other Observers, and Corresponding Group Indices, by Film, Observation and Category of Behavior (Series A) *

Standard Observer	Co-Observers	Observation	Material							Self							Person						
			Films							Films							Films						
			I	II	III	IV	V	VI	I–VI	I	II	III	IV	V	VI	I–VI	I	II	III	IV	V	VI	I–VI
W	XYZ	1	6.0	14.5	5.4	4.1	8.7	7.6	8.3	3.1	2.0	0.6	1.3	1.0	2.2	1.7	1.8	3.0	2.1	1.8	3.0	7.6	3.4
		2	3.7	6.3	3.1	6.3	5.0	9.1	6.4	1.5	1.6	0.7	0.8	1.1	1.8	1.3	2.0	2.3	3.4	1.7	3.4	4.1	2.7
		3	4.7	3.6	3.6	2.8	1.5	3.3	3.1	1.2	1.0	0.5	0.8	0.8	1.4	1.0	4.0	2.4	2.3	0.9	3.0	3.2	2.5
		Av.	4.7	7.8	3.9	4.4	4.9	6.6	5.8	1.9	1.6	0.6	0.9	1.0	1.8	1.3	2.6	2.6	2.6	1.5	3.1	4.9	2.8
X	WYZ	1	5.7	6.9	7.5	3.2	5.7	13.6	7.7	3.1	3.3	0.6	1.1	1.4	1.9	1.9	2.4	2.3	1.6	2.5	3.6	5.9	3.1
		2	2.1	3.3	3.8	3.6	3.7	4.9	3.8	1.9	2.1	0.6	1.2	1.2	1.9	1.5	2.9	3.6	1.7	2.0	4.4	4.7	3.3
		3	6.1	1.7	1.8	2.6	3.6	4.0	3.5	1.9	1.8	0.5	0.5	0.5	1.5	1.1	2.2	1.7	1.3	2.6	4.0	5.0	2.8
		Av.	4.6	3.7	4.3	3.1	4.3	7.0	4.8	2.3	2.4	0.5	0.9	1.0	1.7	1.5	2.5	2.5	1.5	2.4	4.0	5.2	3.1
Y	WXZ	1	15.9	11.5	2.7	5.8	5.3	7.0	8.5	1.2	1.3	0.6	0.7	1.3	3.3	1.4	8.1	4.6	1.6	2.9	10.5	7.3	5.4
		2	7.5	5.2	6.0	7.4	5.1	3.1	5.4	0.8	1.7	0.6	0.7	1.1	2.6	1.3	4.0	2.7	1.0	2.7	8.0	6.6	4.0
		3	6.7	6.5	4.2	2.8	2.3	4.9	4.6	0.5	0.6	0.4	0.3	0.8	1.9	0.8	4.6	3.2	2.0	2.4	4.7	3.6	3.3
		Av.	9.9	7.5	4.4	5.3	4.1	4.7	6.0	0.8	1.2	0.5	0.6	1.1	2.6	1.2	5.5	3.5	1.5	2.7	7.7	5.8	4.2
Z	WXY	1	9.7	8.8	9.5	6.2	8.2	11.0	8.9	2.0	1.6	0.7	0.9	0.9	1.6	1.2	9.5	6.5	1.4	3.4	7.2	10.1	6.5
		2	5.2	5.6	0.9	4.1	7.6	4.2	5.0	1.1	1.7	0.4	1.0	0.6	1.8	0.9	5.2	6.6	3.4	3.5	8.5	5.9	5.4
		3	3.4	4.9	5.0	4.9	4.5	2.8	4.0	1.6	1.1	0.2	0.5	0.8	1.4	0.9	3.4	3.2	1.7	1.2	3.7	6.3	3.3
		Av.	6.0	6.2	5.1	5.1	6.6	5.6	5.8	1.6	1.2	0.4	0.8	0.8	1.6	1.0	6.0	5.4	2.2	2.7	6.5	7.4	5.1
Group		1	10.0	10.8	6.3	4.9	7.0	10.1	8.4	2.4	2.1	0.6	1.0	1.2	2.2	1.6	5.7	4.2	1.7	2.6	6.2	7.8	4.6
		2	4.8	5.2	3.6	5.4	5.4	5.4	5.2	1.3	1.6	0.6	0.9	1.0	2.0	1.3	3.6	3.9	2.4	2.5	6.2	5.4	3.9
		3	5.3	4.3	3.7	3.3	3.0	3.7	3.8	1.3	1.1	0.4	0.5	0.7	1.6	0.9	3.6	2.6	1.8	1.8	3.9	4.6	3.0
		Av.	6.5	6.4	4.4	4.5	5.0	6.0	5.6	1.7	1.6	0.5	0.8	1.0	1.9	1.3	4.2	3.6	2.0	2.3	5.4	5.9	3.8

* All units weighted by number of persons recording.

class 0.5–1.0%, and range from 0.1 to 4.6%, with the median at 1.0%. The indices for the *Person* category are intermediate in

Histogram Series IV. Distribution of 216 Pairwise Percentage Indices of Interpretive Disagreement by Category, Series A.



size between those for *Self* and for *Material*, with a range from 0 to 12.4% and the median at 3.2%. It appears, from these data, that observers interpret self behavior with relatively high agree-

ment and that they tend to identify behavior involving persons with greater uniformity than behavior involving materials.

It is of some interest to see whether these differences between categories persist when allowance is made for chance fluctuations. The differences between *Material* and *Self*, as shown in the following summary, are consistently significant, the *Material* indices of disagreement being always greater than the *Self* indices by an amount at least three times the standard deviation of the differences. The *Material-Person* differences, however, are probably significant for the first four films only, those in the last two films being probably due to chance factors alone.

Differences and 3 σ Differences between Interpretation Indices for Categories Three Observations Combined, Series A, by Films, All Observers

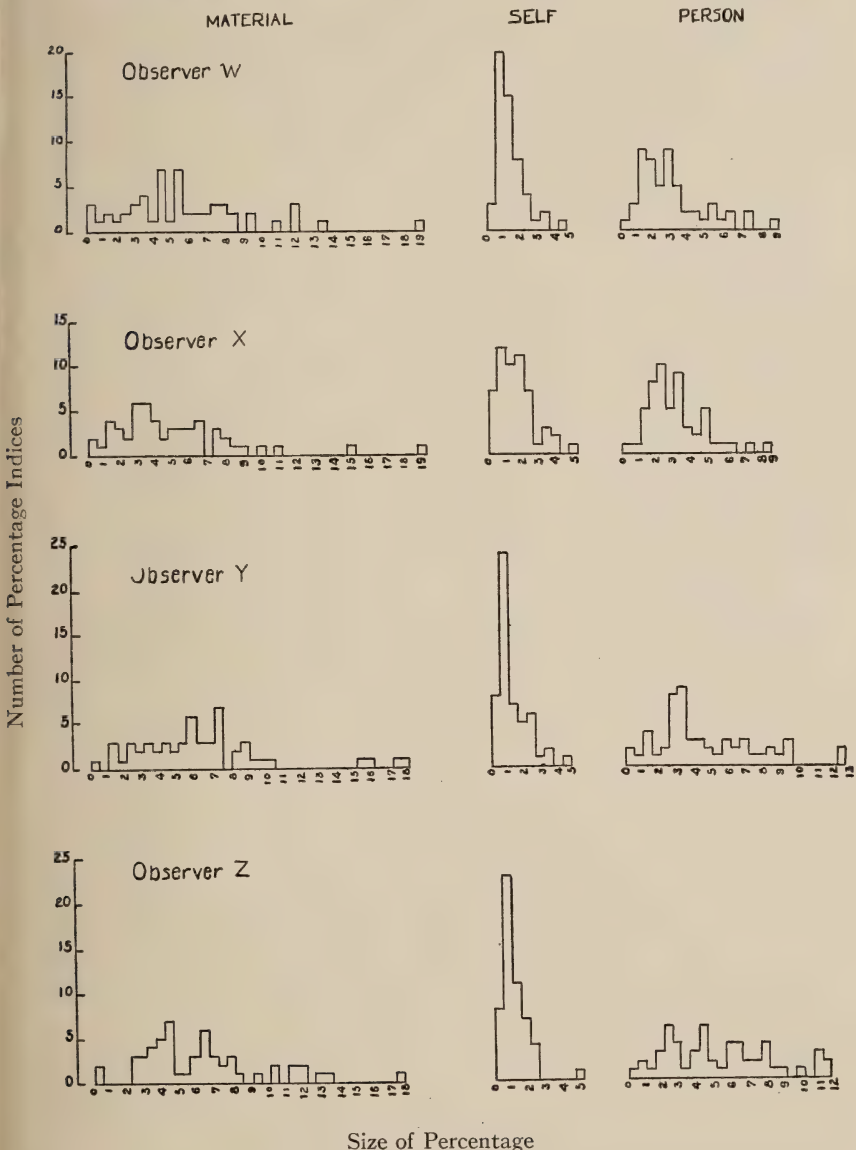
Paired Categories	Film	I	II	III	IV	V	VI	Total
Material-Self	Diff.	+4.8	+4.8	+3.9	+3.7	+4.0	+4.1	+4.3
	3σ Diff.	1.0	0.9	1.2	0.7	0.7	0.7	0.4
Material-Person	Diff.	+2.3	+2.8	+2.4	+2.2	−0.4	+0.1	+1.8
	3σ Diff.	1.2	1.0	1.3	0.8	1.1	0.9	0.4
Self-Person	Diff.	−2.5	−2.0	−1.5	−1.5	−4.4	−4.0	−2.5
	3σ Diff.	0.7	0.4	0.5	0.4	0.9	0.6	0.3

This analysis confirms our other evidence that in this series of observations factors are present which tend to produce, for all observers, the highest amount of disagreement in the interpretation of the *Material* category and the least in the *Self* category.

Extent of Interpretive Disagreement per Observer

To what extent do observers differ in the amount and variability of their interpretive disagreement? Does an observer tend to record one aspect of behavior with greater or less disagreement than his co-observers? Individual differences among observers in the distribution of pairwise indices by category are indicated in Histogram Series V. The first set of figures represents the distribution by category of indices for Observer W, paired with each other observer, in each observation of each of the six films of Series A. The second group of figures shows the corresponding distributions for Observer X, the third, for Observer Y, the fourth, for Observer Z. The upper limit of the range of pairwise indices

Histogram Series V. Distribution of 54 Pairwise Percentage Indices of Interpretive Disagreement per Observer and Category, Series A.



of interpretive disagreement for the *Material* category is 18.9% for Observer W, 18.8% for X, 17.8% for Y, and 17.3% for Z. As a whole, the indices for the individual observers cluster fairly closely around the central point, as was the case for the group. Y has the highest median, 5.8%, and X, the lowest. In the *Self* category, not only is the median percentage of disagreement low for each observer but the range is narrow and practically the same for all observers. The lowest index for any observer is 0.1%, the highest, 4.6%. X has the highest median index of disagreement, 1.4%; Z, the lowest, 0.9%. In the *Person* category, Y and Z have the widest ranges in their indices, from 0 to 12.4% and 0.4 to 11.1%, respectively. Indices for Observers W and X range from 0.4% to 8.7% and from 0 to 8.3%, respectively. Observer W has the lowest median index of disagreement in this category, at 2.6% and Observer Z, the highest, at 4.5%.

The medians of all indices of interpretive disagreement for Series A (all films and observations combined) are shown below by category for each observer and for the group. For comparative purposes, the mean index of disagreement (*not* the mean of the indices) is presented. It will be noted that in all cases, the

Mean and Median Index of Interpretive Disagreement by Observer and Category, All Observations and Films, Series A.

Category	Observers								All Observers	
	W		X		Y		Z			
	M	Md	M	Md	M	Md	M	Md	M	Md
Material	5.8	5.1	4.8	3.9	6.0	5.8	5.8	5.7	5.6	5.2
Self	1.3	1.1	1.5	1.4	1.2	0.9	1.0	0.9	1.3	1.0
Person	2.8	2.6	3.1	2.7	4.2	3.5	5.1	4.5	3.8	3.2

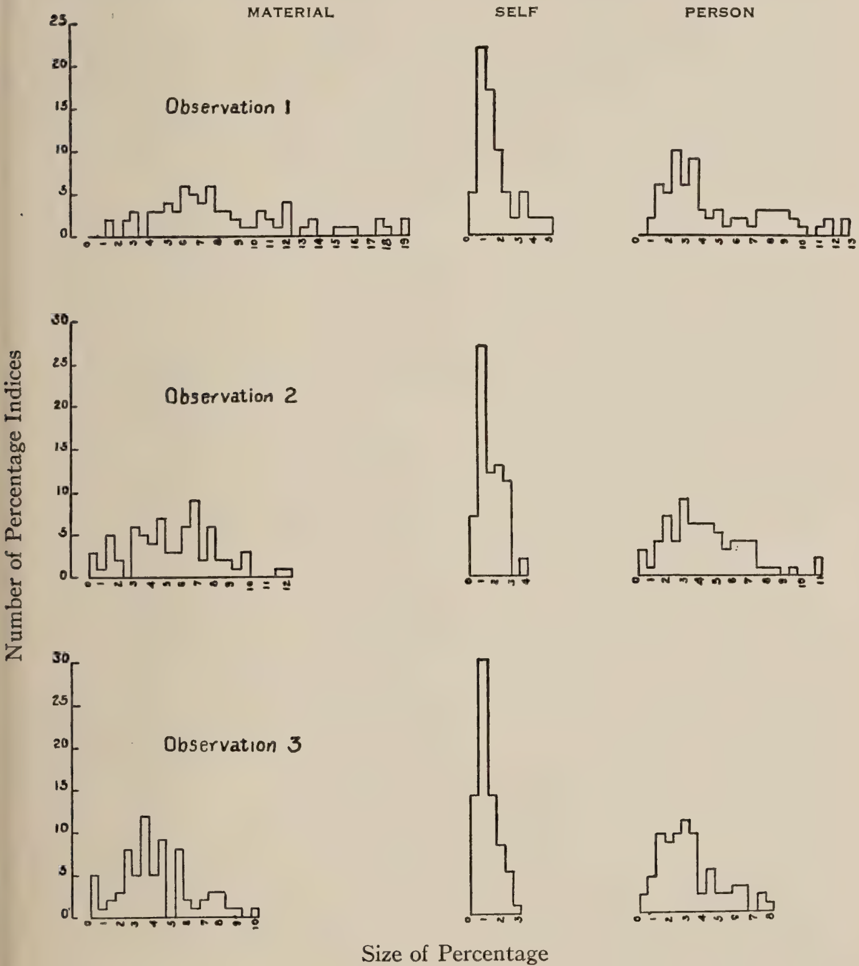
median index and the mean index are close together but that the median index tends to be somewhat smaller and probably more stable than the mean. As to the individual differences among observers, X's average and median index in *Material* are probably significantly lower and Z's indices in *Person* are probably significantly higher than those for the group.

Improvement with Repeated Observation of the Same Behavior

Histogram Series VI shows the effect of repeated observation upon agreement in interpretation by category (all observers

combined). In the *Material* category, the range, the mean, and the median decrease with each successive observation. In the

Histogram Series VI. Distribution of 72 Pairwise Percentage Indices of Interpretive Disagreement by Category and Observation, Series A.



first observation, the range is 1.3–18.9%, in the second, 0–11.9% and in the third, 0–9.6%. The means decrease progressively from 8.4 to 5.2 to 3.8%, the medians, from 7.1 to 5.0 to 3.5% (Table XVI). Even in the *Self* category where the indices are

low throughout and cluster strongly about the central tendencies the decrease with repeated observations is clear-cut. The upper limit of the range in this category decreases from 4.6 to 3.2 to 2.6% from the first to the third observation, the lower limit being 0.1% for first and third observations and 0.3% for the second. The means decrease from 1.6 to 1.3 to 0.9%, the medians, from 1.3 to 1.1 to 0.9%. In the *Person* category, the upper limit for the three successive observations decreases from 12.4 to 10.8 to 7.5%. The means for the first, second and third observations, respectively, are 4.6%, 3.9% and 3.0%. The median is higher in the second observation than in the first, 3.7% as compared with 3.4%, but falls to 2.7% in the third observation.

The means and medians of the indices of interpretive disagreement, by observer, category and observation, are shown in the following table:

Mean and Median Index of Interpretive Disagreement for Each Observer and for the Group in the First, Second and Third Observations of Series A. (All Films Combined)

Category	Observation	Observers								All Observers	
		W		X		Y		Z			
		M	Md	M	Md	M	Md	M	Md	M	Md
Material	1	8.3	6.8	7.7	6.0	8.5	7.3	8.9	8.0	8.4	7.1
	2	6.4	5.5	3.8	3.2	5.4	6.0	5.0	4.2	5.2	5.0
	3	3.1	3.3	3.5	3.1	4.6	4.8	4.0	4.0	3.8	3.5
Self	1	1.7	1.3	1.9	1.6	1.4	1.0	1.2	1.0	1.6	1.3
	2	1.3	1.2	1.5	1.5	1.3	1.0	0.9	0.8	1.3	1.1
	3	1.0	0.9	1.1	0.9	0.8	0.7	0.9	1.0	0.9	0.9
Person	1	3.4	2.7	3.1	2.8	5.4	5.2	6.5	7.5	4.6	3.4
	2	2.7	2.7	3.3	3.5	4.0	3.5	5.4	5.8	3.9	3.7
	3	2.5	2.2	2.8	2.2	3.3	3.2	3.3	2.8	3.0	2.7

We have seen that all observers improve in interpretation of behavior from the first to the third observation. What is the general pattern of improvement by category and observation for the group and for individual observers? Do observers improve in some categories at the expense of others, or does improvement occur regularly in all categories in each successive observation?

To throw light on the individual differences in improvement, we turned to the basic data on agreements and disagreements given in Table XV. Here we see that Observer W disagrees with his co-observers in the interpretation of *Material* in 241 units in the first observation, 229 in the second and 114 in the third.

Thus there was a 5.0% decrease from the first to the second (12/241) and a 50.2% decrease from the second to the third (115/229). At the same time the number of agreeing units (total recorded units weighted) increased from 2895 to 3564 to 3678, or 23.1% (669/2895) from the first to the second, and 3.2% (114/3564) from the second to the third. The net effect of these changes was to decrease the index of disagreement from 8.3 to 6.4 to 3.1%. The percentage increases in agreement and the percentage decreases in disagreement, for all films of Series A, computed in this manner, are shown below by category and observer, together with the corresponding figures for the group.

CATEGORY	OBSERVATION	PERCENTAGE INCREASE IN AGREEMENTS					PERCENTAGE DECREASE IN DISAGREEMENTS				
		Observers				All Observers	Observers				All Observers
	From–To	W	X	Y	Z		W	X	Y	Z	
Material	1–2	23.1	15.6	19.3	13.8	17.9	5.0	43.4	24.3	36.0	27.1
	2–3	3.2	7.0	6.5	5.9	5.6	50.2	2.4	8.9	14.8	22.4
Self	1–2	0.4	0.5*	1.2	1.7	0.7	23.0	19.8	6.1	23.1	18.1
	2–3	1.9	1.9	1.5	2.8	2.0	22.1	25.9	40.4	2.9†	23.6
Person	1–2	3.5	0.7	1.1	3.5	2.2	18.9	4.8†	24.9	13.1	14.7
	2–3	2.3	3.1	0.7	3.0*	0.7	5.5	11.7	16.0	40.6	22.4

* Decrease † Increase

In the recording of *Material*, all four observers show the same general tendency with respect to proportionate increase in agreement, i. e., a much higher percentage increase between the first and second observations than between the second and third. In disagreements, W differs distinctly from his three co-observers. While X, Y and Z show decreases in disagreements amounting to 43%, 24% and 36%, respectively, between the first and second observations, Observer W has only a 5% decrease between first and second observations but a 50% decrease between the second and third, compared with 2%, 9% and 15% respectively for X, Y and Z. In the *Self* category, all observers made a slightly greater percentage improvement in agreements between the second and third observations than between the first and second. In disagreements, however, there are marked individual differences. W shows high percentage improvement, almost evenly balanced between the two repetitions; X shows marked improvement, with a higher percentage in the second repetition; Y shows

practically all of his improvement between the second and third observations; and Z shows his improvement wholly between the first and second observations, having an actual increase in disagreements between the second and third observations. In number of agreements in *Person*, Observers W and Y have a slightly larger percentage increase between the first and second observations than between the second and third. The reverse is true for X, and Z has a 3.0% decrease in number of agreements in the third observation. W and Y have a greater decrease in disagreements in the first repeated observation than in the second, Z reverses this order and X has a 4.8% increase in the second observation.

To summarize, all observers in recording *Material* and all except X in recording *Person* have the larger percentage increase in agreements between the first and second observations. This order is reversed by all observers in the *Self* category, i. e., the greater percentage increase occurs between the second and third observations. The larger percentage decrease in disagreements occurs between the first and second observations for all observers except W in recording *Material*, for W and Z in recording *Self* and for W and Y in recording *Person*.

As a check on the tendency toward improvement, we made two further tabulations, comparable to those in the preceding chapter. We first tabulated the number of indices which showed "improvement," i. e., where the index of disagreement diminished in size, from one observation to another. We then attempted to estimate the probable significance of the total decreases found. The following table shows the first of these tabulations, by observer and category. It will be remembered that each observer had 18 chances to improve from one observation to another in each category (i. e., in pairwise combination with each of three co-observers in six films).

The tendency toward improvement is most marked in the indices of interpretive disagreement in the *Material* category. There was an average improvement of all observers in 81% (58/72) of the indices from the first to the second observation, and in 60% (43/72) from the second to the third, with an improvement from the first to the third in no less than 92% (66/72); in the *Self* category, the comparable percentages are somewhat less,

61% (44/72) from the first to the second, 64% (46/72) from the second to the third, and 83% (60/72) from the first to the third; in the *Person* category, 62% (45/72) from the first to the second, 65% (47/72) from the second to the third, and 72% (52/72) from the first to the third. These tendencies toward improvement are generally more marked in this tabulation than in the comparable tabulation indicating changes in timing tendencies.[1] If individual tendencies can be indicated by such a crude analy-

Category	Observer	Number of Chances for Improvement	Index of Disagreement less on Second Observation than on First	Index of Disagreement less on Third Observation than on Second	Index of Disagreement less on Third Observation than on First
Material	W	18	13	13	17
	X	18	16	10	16
	Y	18	12	9	16
	Z	18	17	11	17
	All	72	58	43	66
Self	W	18	10	10	14
	X	18	11	11	15
	Y	18	8	14	17
	Z	18	15	11	14
	All	72	44	46	60
Person	W	18	11	13	10
	X	18	8	9	11
	Y	18	15	9	14
	Z	18	11	16	17
	All	72	45	47	52

sis, it seems that X and Z improve most markedly in the early observation of *Material;* that Y improves late and Z early in *Self,* and that Y improves early and Z late in *Person.*

Turning now to an analysis of the probable significance of these changes we compared the differences between the percentage indices for each observer and for the group in the first and in the second observations and the differences between percentages for the second and for the third observations (Table XVII) with the standard deviations of these differences multiplied by 3. If these differences are greater than three times the deviations to be expected from the operation of chance, they at least suggest a trend toward real improvement. They will also throw light on the possible differences among observers as to pattern of improvement in different categories. These differences between

[1] See p. 147, *supra.*

percentages of interpretive disagreement for the first and second observations and for the second and third observations for Series A (films combined) are shown below by category and observer, together with figures representing the standard deviations of the differences multiplied by 3.

Decreases in Indices of Disagreement on Successive Observations, All Films Combined, Series A, by Observer and Category

Category		From First to Second Observation					From Second to Third Observation				
		Observers					Observers				
		W	X	Y	Z	All	W	X	Y	Z	All
Material	Diff.	1.9	3.9	3.1	3.9	3.2	3.3	0.3	0.8	1.0	1.4
	3 σ Diff.	1.9	1.7	1.9	1.9	1.1	1.5	1.3	1.5	1.5	0.7
Self	Diff.	0.4	0.4	0.1	0.3	0.3	0.3	0.4	0.5	0	0.4
	3 σ Diff.	0.3	0.4	0.3	0.0	0.0	0.0	0.3	0.0	0.0	0.0
Person	Diff.	0.7	0.2*	1.4	1.1	0.7	0.2	0.5	0.7	2.1	0.9
	3 σ Diff.	0.8	0.8	1.1	1.3	0.4	0.8	0.8	0.8	1.1	0.4

* Increase

The group figures in this table indicate that, in the *Person* and the *Self* categories, the differences between indices for successive observations are consistently in the direction of improvement and are more than three times their standard deviations. They are not large, however, and suggest approximately the same amount of improvement in the second, as in the first, repeated observation. There seems to be a well defined trend toward greater agreement among these observers in the interpretation of *Person* and of *Self* in each repetition. In the *Material* category, however, the situation differs in two respects. Here the improvement is more clearly significant than in either of the other categories, as indicated by the excess of the differences over three times the standard deviations of the differences. There is also an indication that the improvement from the first to the second observation is greater than that from the second to the third. This latter point we explored in some detail to determine whether there are any group or individual trends toward a particular pattern of change in successive observations of material behavior. The greater improvement in the recording of *Material* than of other categories may be affected by three factors. Since *Material* was the preferred category, changes in interpretation might be expected particularly to affect this cate-

gory; since, in this series, material behavior occurred least frequently, a small change in absolute number of disagreements would be reflected in a relatively large change in percentage; and it is also possible that the larger the amount of disagreement at first (the indices in *Material* being larger than those in *Self* and *Person*), the more probable is its subsequent decrease. This last point is exemplified by the differences in improvement in the several films. The largest changes in indices occurred in those films having the largest amount of disagreement in the earlier observation, the reverse also being true. In three films, group indices of 10.0, 10.8 and 10.1% were lowered to 4.8, 5.2 and 5.4%, respectively, in the subsequent observation. In Film IV, with the lowest amount of disagreement on the first observation, 4.9%, there was even a slight, though insignificant, increase in the following observations (5.4%).

Considering the individual observers separately, we find that in the first repetition W's improvement in the *Material* category is barely significant according to our criterion, whereas the improvement of the other three observers is distinctly greater than that attributable to chance. But in the second repetition W shows marked improvement while the change in the indices of each other observer is insignificant. In *Self*, Y's improvement is insignificant from the first to the second, but clearly significant from the second to the third observation; and in *Person*, Z's improvement is insignificant from the first to the second, and clearly significant from the second to the third observation. Thus the individual patterns suggested in the preceding analysis are confirmed in X's and Z's early improvement in *Material*, Y's late improvement in *Self* (but not Z's early improvement) and Z's late (but not Y's early) improvement in *Person*.

A more detailed and complete discussion of the individual interpretation biases of the observers and of the situational factor as an element producing differential interpretation indices will be found in the two following chapters. Two further points may be explored, in a limited way, in connection with the data presented in this chapter, i. e., questions relating to the variation between different pairs of observers and questions as to the total effect of an extremely deviating observer. The first of these points has definite implications as to the adequacy of a pairwise comparison in determining reliability coefficients.

Variability of Pair Indices Compared with Corresponding Group Indices

Assuming that an event is more probable if a larger number of equally trained and capable persons record it, what is the relation of this more probable index to that of each of the separate pairs comprising it? For any given observation, some pairs will probably show positive deviations from the combination of all pairs (i. e., greater disagreement) and some will show negative deviations (i. e., less disagreement). A fundamental question is whether, in a series of samples, the deviations for specific pairs are consistently positive or negative. If we conceive of a pairwise index as representing the position of any two observers in a distribution of many observers, would a correction for the variability of this particular pair be ± (i. e., would they deviate from the norm in both directions) or would it be either + or − (i. e., would this pair be placed consistently either above or below the norm of the distribution)?

Our data, of course, are not suitable for solving this problem. We can merely consider our six pairwise combinations in relation to the "group," using for this purpose our most adequate series, the six films in Series A. Here we have two indices of interpretive disagreement for each observer, the pairwise index representing the relation of each observer to each other observer, and the group index representing the relation of each observer to the other three observers.

It will be noted that the group index from which the deviation of the pair index is computed is calculated for a group containing the paired observer (e. g., W-X and W-XYZ). The presence of the paired observer in the group will naturally tend to bring the observer's index with the group closer to his pair index in proportion to the amount of identical data represented in the two indices. Stated otherwise, the difference between an observer's indices in a given pair and in the group containing the paired observer is too low an index of his differences with different observers, but, with the group as a standard, it is the figure we need for measuring the inaccuracy of the pair index.

For each observer we computed three pair indices and a group index of interpretive disagreement by category, observation and

film. Then, for each observer in each pairing (e. g., for X in the pairs X-W, X-Y and X-Z; for W in the pairs W-X, W-Y and W-Z, etc.), with his group index as a standard, we found the divergence of his pair index from his group index and recorded it as + if his pairwise index of disagreement were greater than his index with the group as standard, and as − if it were less. We then made a distribution of the deviations of each observer's pair indices from the corresponding group index by category and observation, using a step interval of 0.5%. These deviations are shown in Table XVIII. Looking at these deviations, first from the standpoint of the category, we see that pair and group indices for each observer approach each other most closely in the *Self* category and scatter most widely in the *Material* category. When the group index is taken as the standard of reliability, pair indices are least reliable, i. e., they diverge most widely, in the first observation of the *Material* category. The range here is from + 6.3% (2 cases) to − 5.7 and − 5.6%. This suggests the need for considerable caution in accepting the percentage of agreement in interpretation of a single pair of observers as the measure of reliability of a technique. A more frequent fallacy, perhaps, to which this wide range of deviation of pair indices calls attention, is the assumption that the agreement of two observers is proof of occurrence of a particular item of behavior.

We are interested in these deviations of an observer's pair indices from his group index when their size is greater than can be ascribed to chance factors and also when a given pair of observers exhibits a constant negative or positive bias. Even small deviations may be considered of some significance when their direction is so unvarying as to suggest the operation of a constant factor. The pair indices of one observer in two pairings in one category deviated from the group index in a constant direction in every observation of every film, with the exception of one pairing in one observation, that is, 17 and 18 times, respectively, when there were 18 chances for variation. Observer W in the recording of *Self* when paired with X shows a consistent tendency to have a lower index of agreement than when paired with the group, the range of differences being from 0 to − 2.2%. On the contrary, W in the WZ combination has, with the exception of one film-observation, a slight positive deviation, or in-

TABLE XVIII

Distribution of the Deviations of All Pairwise Indices of Interpretive Disagreement from the Corresponding Group Indices as Standard for Observers W, X, Y, and Z, by Category, Observation and Pairing of Observers (Series A).

SO = Standard Observer . CO = Co-Observer

OBSERVATION 1

Category	SO	CO	Deviations																										
			−6.5	−6.0	−5.5	−5.0	−4.5	−4.0	−3.5	−3.0	−2.5	−2.0	−1.5	−1.0	−0.5	0	+0.5	+1.0	+1.5	+2.0	+2.5	+3.0	+3.5	+4.0	+4.5	+5.0	+5.5	+6.0	+6.5
Material	W	X												1		1		2		1					1				
		Y							1		1							1	1				2						
		Z								1	1	1	1		1								1						
	X	W									1	1		1				2	1										
		Y											1			1	1			1					1		1		
		Z			1									1	2		1						1						
	Y	W						1						2	1				1			1							
		X											1			½	½		3										1
		Z							1					1		1	2		1										
	Z	W										1		1	1					2				1					
		X					1								1	1		1				1	1						
		Y			1						1		2										1						1
Self	W	X										1			3	2													
		Y												1		1½	2½		1										
		Z														½	3½	1	1										
	X	W										1			1	4													
		Y														4	1			1									
		Z															2	4											
	Y	W												1	1	2	2												
		X													2	3½	½												
		Z														1	1	2	1	1									
	Z	W												1		3	2												
		X												1	1	2	2												
		Y														2	3					1							
Person	W	X												1	1			2	1		1								
		Y											1		2	½	2½												
		Z													2	2		1	1										
	X	W													1	2	2		1										
		Y													1	1½	1½	1			1								
		Z									1					3		2											
	Y	W								1					1		2	1		1									
		X														1		2	1	1	1								
		Z							1				1	1		½	1½		1										
	Z	W															2	3		1									
		X											1			½	1½	1	2										
		Y										1	2	1	1			1											

TABLE XVIII—(*Continued*)

Distribution of the Deviations of All Pairwise Indices of Interpretive Disagreement from the Corresponding Group Indices as Standard for Observers W, X, Y, and Z, by Category, Observation and Pairing of Observers (Series A).

SO = Standard Observer CO = Co-Observer

OBSERVATION 2

Category	SO	CO	−6.5	−6.0	−5.5	−5.0	−4.5	−4.0	−3.5	−3.0	−2.5	−2.0	−1.5	−1.0	−0.5	0	+0.5	+1.0	+1.5	+2.0	+2.5	+3.0	+3.5	+4.0	+4.5	+5.0	+5.5	+6.0	+6.5
Material	W	X										1	1	1				1		1	1								
		Y								1		2				1						1			1				
		Z											1		1			1	2	1									
	X	W										3			1	1		1											
		Y											1	1		½	1½				1				1				
		Z											2					1			2	1							
	Y	W									1		1	1	1			1	1										
		X											1			1	1		1	2									
		Z													1		2	1	2										
	Z	W										1	2	1	1				1										
		X											1			1	1				2			1					
		Y											1	2	1					1				1					
Self	W	X													2	4													
		Y														1½	4½												
		Z														½	4½	1											
	X	W													3	1	2												
		Y														3½	2½												
		Z														2	2	1	1										
	Y	W													1	2	2	1											
		X														5	1												
		Z														2	2	2											
	Z	W													1	3	2												
		X													2	3	1												
		Y														2½	2½		1										
Person	W	X													1	1½	2½			1									
		Y										1			1	1		1	1		1								
		Z											1	3		½	½					1							
	X	W													1	1	1		2	1									
		Y													2	1	1	1	1										
		Z											2	1	1			1	1										
	Y	W													1	1	1	1	1		1								
		X												1		1	1	1	2										
		Z											1	2	1		2												
	Z	W												1			3				2								
		X											1			1½	1½		1		1								
		Y						1							2	2	1												

TABLE XVIII—(*Continued*)

Distribution of the Deviations of All Pairwise Indices of Interpretive Disagreement from the Corresponding Group Indices as Standard for Observers W, X, Y, and Z, by Category, Observation and Pairing of Observers (Series A).

SO = Standard Observer CO = Co-Observer

OBSERVATION 3

Category	SO	CO	Deviations																										
			−6.5	−6.0	−5.5	−5.0	−4.5	−4.0	−3.5	−3.0	−2.5	−2.0	−1.5	−1.0	−0.5	0	+0.5	+1.0	+1.5	+2.0	+2.5	+3.0	+3.5	+4.0	+4.5	+5.0	+5.5	+6.0	+6.5
	W	X												1	1			1		1	1	1							
		Y									1			2			1	1		1									
		Z									1			1		½	½	3											
	X	W											1	1		1		2	1										
Material		Y									1		1					2	1	1									
		Z											2			1	1			1	1								
	Y	W							1	1					2	1				1									
		X															1	1		2			1	1					
		Z								1	1				1		1	1		1									
	Z	W												1	3			1				1							
		X													2		2			1		1							
		Y									1		1	1	1				1	1									
	W	X													1	4½	½												
		Y														2	3	1											
		Z														1½	4½												
	X	W													2	3½	½												
Self		Y														1	3	2											
		Z													1	1	3	1											
	Y	W													1	2½	2½												
		X													1	1½	3½												
		Z														½	5½												
	Z	W														3½	2½												
		X														3	3												
		Y														2½	2½	1											
	W	X											1		2	1	1				1								
		Y									1				3				2										
		Z													2	½	1½	1				1							
	X	W													1	4		1											
Person		Y										1		1	2			1			1								
		Z											1		1		1	1	1				1						
	Y	W														½	1½	3	1										
		X							1					1				3	1										
		Z											2		2			1			1								
	Z	W													2	1	1	1	1										
		X													1	2	1		1		1								
		Y												2	2			1	1										

creased disagreement, from − 0.2% to + 1.2%. It seems that W tends consistently by a slight margin, varying from 0 to 2.2%, to agree more with X in the interpretation of *Self* than with the group, while, by a margin ranging up to 1.2%, he tends to agree less with Z than with the group.

The conclusion we draw from this consideration of the relation between observers' pair and group indices is methodological. The practical limitation in life situations of the number of simultaneous observers makes the relative reliability of a pair index an important problem.

Effect of one Observer's Record on Indices of Other Observers

The question as to the effect of the deviating observer must also be approached in a very cursory manner. It has been emphasized that our indices measure relative rather than absolute reliability. The probable occurrence of events is inferred from the extent of agreement among observers and agreement with co-observers is taken as the test of reliability. Obviously, the records of each observer will affect the interpretive "error" of all other observers. Since the base of our index of interpretive disagreement is the total number of weighted units of a given category recorded by one observer, there can be no disagreements from the standpoint of this category if the observer records no behavior in the category. The disagreements will be found in the substituted category. If one observer out of four differs from the other three by recording less of a given kind of behavior, he may thereby reduce the number of his disagreements in this category and, consequently, his index of unreliability. For this category, the three observers who agree in recording more of the particular behavior have higher totals of occurrence, larger group indices of disagreement and larger pair indices for each pair in which the deviating observer is included. Compensation occurrs, of course, in the category or categories which the deviant observer has recorded in place of the category he has underemphasized. The observer who differs from the others will have more disagreements and a higher index here than his co-observers. This kind of situation, in which one observer raises the indices of disagreement of three others who agree in

opposition to him, becomes numerically significant when the behavior which the differing observer neglects or misclassifies itself occurs infrequently, so that relatively few disagreeing units have an appreciable effect upon the percentage index.

We will merely present one example to indicate that exactly the situation described above does occur, i. e., that the three observers who agree on the occurrence of a large amount of behavior in a certain category have higher indices of disagreement in that category than the one observer who differs from them. The illustration is taken from Observation I of Film VI in Series A, in which Y recorded a relatively small amount of behavior in the *Material* category, as compared with the other observers.

*Number of Units Recorded**

Observer	Category	
	Material	Self
Y	(1) 113	(4) 913
W	145	892
X	162	883
Z	164	806

Indices of Disagreement

Standard Observer	Co-Observers	Material	Self
Y	W	(2) 5.7%	(5) 2.4%
W	Y	10.6%	1.0%
Y	X	(2) 8.2%	(5) 3.2%
X	Y	18.8%	1.7%
Y	Z	(2) 7.3%	(5) 4.3%
Z	Y	17.3%	1.8%
Y	WXZ	(3) 7.0%	(6) 3.3%
W	XYZ	7.6%	2.2%
X	WYZ	13.6%	1.9%
Z	WXY	11.0%	1.6%

* These units are unweighted.

The relationships in detail are as follows:

(1) Observer Y records fewer units in the *Material* category than are recorded by W, X and Z.

(2) Y has decidedly lower pair indices in this category than his co-observers when they are paired with him.

(3) Y has a lower group index in this category than the other observers.

(4) Y compensates for his underemphasis of *Material* by recording more *Self*.
(5) In the *Self* category Y has a higher measure of disagreement than the other observers paired with him.
(6) Y's group index in *Self* is higher than those of the other observers.

The opposite effect appears when the deviating observer classifies a much larger proportion of behavior in a certain category than do the other observers. He then tends to have a high index of disagreement in this category and a correspondingly lower one in the category in which he has recorded less behavior than his associates. Observer Y in Film I, Observation 1, may be used as an illustration. From the basic data in Tables XV A and B, we see that Y recorded 26.4% more "units of agreement" of *Material* (503 total weighted units) than the group average (398) and 4.5% less of *Self* (2781 total weighted units compared with the group average of 2913). The difference in size of base, of course, accounts in part for the difference in percentages. Y's index of disagreement in *Material* (Table XVII) is 15.9%, as compared with 6.0%, 5.7% and 9.7% for the three co-observers and his index in the *Self* category is 1.2%, while the indices for the others are 3.1%, 3.1% and 2.0%.

The biases of individual observers as shown in the tendency to record more or less of a given category constitute one of the several factors that influence the measures of interpretive disagreement. When a bias toward underemphasis of a category is clearly established for a given observer, we may expect a lowering of his index of disagreement. If, on the other hand, he tends to record more of a certain kind of behavior than do his associates, we may expect his index of interpretive disagreement to be relatively larger. Individual biases will be considered further in the next chapter.

Measurement of Interpretive Disagreement with Partial Allowance for Timing Discrepancies

It is highly probable that a fair proportion of the interpretation disagreement is spurious, i. e., produced by the differential timing tendencies of the observers rather than by their inability to agree on the classification of the behavior observed. The indices of reliability which we used in the nursery school and

kindergarten studies made an allowance for timing misplacement by counting as agreements those units[1] in which one observer of a pair agreed with the other in an adjacent unit. The underlying assumption in this correction was that timing discrepancies operated entirely in the direction of increasing disagreement. Therefore, all units in which the co-observer had a "synchronous" record disagreeing with the standard observer but an agreeing record in an adjacent unit were counted as agreements. These questionable agreements were weighted only 1, whereas the complete agreements ("synchronous" units agreeing) were weighted 2. The effect of such a correction is, of course, to decrease the number of total disagreements, increase the number of total agreements, and so reduce the indices of disagreement. Purely for purposes of comparison, we present below the group indices for each observer in the third observation of Series A (all films totaled) to show the net effect of such a correction.[2]

It is evident from the following table that the net result of

Indices in Third Observations of Films I–VI, Totaled
(1) allowing agreements in adjacent units
(2) crude indices with no such allowance

Standard Observer	Co-Observer	Material (1)	Material (2)	Self (1)	Self (2)	Person (1)	Person (2)
W	XYZ	1.5%	3.1%	0.4%	1.0%	1.4%	2.5%
X	WYZ	1.1%	3.5%	0.4%	1.1%	1.5%	2.8%
Y	WXZ	1.8%	4.6%	0.3%	0.8%	1.6%	3.3%
Z	WXY	1.6%	4.0%	0.4%	0.9%	1.6%	3.3%

this method of allowing for timing discrepancies was to cause a decided decrease in the size of indices. In all categories, the consideration of disagreements adjacent to agreements as agreements reduced the size of the indices to one-half or even one-third the size of the crude indices.

An Attempt at Impartial Correction for Timing Discrepancies

We are by no means satisfied that the index presented in the first section of this chapter is an adequate index of the degree of

[1] Unit is here used in the broad sense of time interval. It will be remembered that the unit in these earlier studies was the five-second interval.

[2] A minor difference between the method used by Arrington and that used in these calculations is our consideration here of each pair of categories separately with no correction for those units counted twice because the co-observer had recorded in them two disagreeing categories. This tends to increase slightly the size of the indices.

interpretive disagreement in our observations. We have used it in our comparisons, to show the differences between categories and observers and the tendency toward change with consecutive observations, and we shall use it again in the two following chapters to indicate individual patterns in disagreements and situational differences. To the extent that we have applied a thoroughly consistent and objective method to all of our data, we may be considered justified in presenting this index in this way. The corrected index discussed briefly in the preceding section, however, seems to us to be logically unjustified because of the underlying assumption that timing tendencies operate always to produce greater numbers of disagreements and because of the failure to take into account possible spurious agreements produced by these same factors. We have, therefore, discarded this type of correction as invalid.

We attempted to develop an automatic correction for timing discrepancies, based on the conception that every unit was likely to be misplaced backwards or forwards. This was applied in the analysis of agreements as well as disagreements.

First, as in the timing tabulation, the record of each observer was weighted for agreement with other observers recording in the preceding and following as well as in the "synchronous" units, so here, the disagreements for each observer were weighted in terms of the disagreeing records of other observers in the preceding and following as well as in the "synchronous" units. Thus, for a given observer, a minimum disagreement occurred for a given unit if he disagreed only with the synchronously recorded units of the other observers. A maximum disagreement occurred if a given unit disagreed not only with the "synchronous" but also with the preceding and following units. Each disagreeing unit, in other words, was given a chance to be weighted as high as 3 or as low as 1; each agreeing unit was likewise given a chance to be weighted as high as 3 or as low as 1. We worked out such indices for Series A, with the surprising result that, although the trend of the two indices was very similar, this "corrected" index was generally larger.[1] Thus, for all

[1] We do not present the details of the construction of this index; it was unsatisfactory in conception and effect (e. g., it introduced accidental weights). It is merely given here to illustrate a stage in our procedure.

observers 83.3% of such indices in the *Material* category were greater than were the crude indices, 58.3% of the indices in the *Self* category and 79.2% of the indices in the *Person* category; or, 73.6% of all the indices were greater, 13.9% were unchanged and only 12.5% were smaller when calculated by this method than were the crude indices.

Thus this attempt to construct an index which would allow for the timing factor indicated that the proportion of *disagreements* rather than of *agreements* tended to increase as a result of this manipulation.

The implications of this result made it clear than an intensive examination of the mechanical factors operating for and against interpretive agreement in a timing technique is more important, at this stage, than the application of any correction determined *a priori*. Chart X represents a schematization of the changes that would occur in various combinations of units if each unit of the observed category recorded by the standard observer is conceived as being misplaced by one unit and as belonging in the unit either preceding or following.

Our next stage in analyzing our data will be to examine in detail the presumable effect of timing errors upon each of these classes of interpretation disagreements, the effect upon our indices of the exclusion of certain classes of partial agreements and disagreements, and the effect upon the indices of the limit of accuracy that we have, up to this point, accepted. At the same time, we shall present, for comparison, a series of observations in which the timing factor has been partially controlled through mechanical aids.

CHART X—POSSIBLE CLASSIFICATION OF RECORDED UNITS OF BEHAVIOR IF PRECEDED OR FOLLOWED BY EACH OF THE FOLLOWING COMBINATIONS, ON THE ASSUMPTION THAT THE STANDARD OBSERVER'S OBSERVED CATEGORY IS SHIFTED ONE UNIT, AND CHANGES IN AGREEMENT RESULTING FROM THIS SHIFTING

SO = Standard Observer   CO = Co-observer

(+) Greater Agreement   (−) Greater Disagreement   (o) No Change

ACTUAL CLASSIFICATION	(1) SO CO	(2) SO CO	(3) SO CO	(4) SO CO	(5) SO CO	(6) SO CO	(7) SO CO	(8) SO CO	(9) SO CO	(10) SO CO	(11) SO CO	(12) SO CO	(13) SO CO	(14) SO CO	(15) SO CO	(16) SO CO
I. Agreement SO CO	I (o)	II (a) (o)	II (b) (o)	III (a) (−)	III (b) (−)	IV (−)	V (a) (−)	V (b) (−)	II (a) (o)	II (b) (o)	I (o)	III (a) (−)	III (b) (−)	IV (−)	V (b) (−)	V (a) (−)
II. Partial Agreement (a) SO CO	I (o)	II (a) (o)	II (b) (o)	III (a) (−)	III (b) (−)	IV (−)	V (a) (−)	V (b) (−)	II (a) (o)	II (b) (o)	I (o)	III (a) (−)	III (b) (−)	IV (−)	V (b) (−)	V (a) (−)
(b) SO CO	I (o)	II (a) (o)	II (b) (o)	III (a) (−)	III (b) (−)	IV (−)	V (a) (−)	V (b) (−)	II (a) (o)	II (b) (o)	I (o)	III (a) (−)	III (b) (−)	IV (−)	V (b) (−)	V (a) (−)
III. Partial Disagreement (a) SO CO	I (+)	II (a) (+)	II (b) (+)	III (a) (o)	III (b) (o)	IV (o)	V (a) (+)	V (b) (+)	II (a) (+)	II (b) (+)	I (+)	III (a) (o)	III (b) (o)	IV (o)	V(b) (+)	V (a) (+)
(b) SO CO	I (+)	II (a) (+)	II (b) (+)	III (a) (o)	III (b) (o)	IV (o)	V (a) (+)	V (b) (+)	II (a) (+)	II (b) (+)	I (+)	III (a) (o)	III (b) (o)	IV (o)	V (b) (+)	V (a) (+
IV. Disagreement SO CO	I (+)	II (a) (+)	II (b) (+)	III (a) (o)	III (b) (o)	IV (o)	V (a) (+)	V (b) (+)	II (a) (+)	II (b) (+)	I (+)	III (a) (o)	III (b) (o)	IV (o)	V (b) (+)	V (a) (+)
V. Uncorroborated Record (a) SO CO	I (+)	II (a) (+)	II (b) (+)	III (a) (−)	III (b) (−)	IV (−)	V (a) (o)	V (b) (o)	II (a) (+)	II (b) (+)	I (+)	III (a) (−)	III (b) (−)	IV (−)	V (b) (o)	V (a) (o)
(b) SO CO	I (+)	II (a) (+)	II (b) (+)	III (a) (−)	III (b) (−)	IV (−)	V (a) (o)	V (b) (o)	II (a) (+)	II (b) (+)	I (+)	III (a) (−)	III (b) (−)	IV (−)	V (b) (o)	V (a) (o)

In this method of analysis, I, II (a) and II (b) are considered agreements; III (a), III (b) and IV are considered disagreements; V (a) and V (b) are considered uncorroborated records.

CHAPTER IX

THE MEASUREMENT OF INDIVIDUAL BIASES IN OBSERVATION

We have been concerned, up to this point, with an analysis of means of determining the degree of error among observers using a given recording technique. We have discussed the variation in the amounts of such errors as we have been able to measure. It is important to know also the extent to which the errors made by each of the observers are biased, i. e., the extent to which they tend to form a pattern which repeats itself in the various relevant situations. In the technique we are analyzing, such biased errors could manifest themselves in systematic timing tendencies and in interpretation errors favoring a given category when the comparison is between one observer and other observers; in a persistence of overemphasis of a given category when the comparison is between the observer's own consecutive records; or in a general overemphasis of some particular category when the comparison is between series of crude observations and an approximation to a "true" record.

Specifically, the questions upon which we attempt to throw light are these:

(1) Where two or more observers are recording simultaneously, do their disagreements with one another in timing events tend to form consistent patterns of retardation or acceleration?

(2) Where two or more observers are recording simultaneously, are their differences in interpretation distributed randomly, or are there consistent individual tendencies for disagreements to favor some one category at the expense of others?

(3) Where one observer is recording the same series of events in several consecutive observations, are his disagreements with himself in a given category randomly distributed or does an overemphasis developing in one part of a given observation tend to reproduce itself throughout the observation, or to form a particular pattern, e. g., of compensation?

(4) When several observers are recording a series of events in

which an approximation to the "true" record of such events is known (a) is there any individual tendency toward overemphasis or underemphasis, and (b) is there a group bias in favor of, or against, a given category?

It will be remembered that the observer differentiated the observed behavior in terms of three categories, representing in brief: whether the character observed was actively manipulating materials (*Material*); whether he was in physical contact with other persons (*Person*); or whether he was neither actively manipulating materials nor engaged in physical contact, but was engaged in what was defined broadly as self activity, such as walking around, talking, etc. (*Self*). Each of these categories was narrowly and arbitrarily limited to certain specified acts and was, by definition, exclusive of the other two categories.

Individual Patterns in Timing Shown in Simultaneous Observations of Different Observers

Each of four observers saw six films in Series A three times and the timing tendency of each (excess of retarded or accelerated units divided by synchronous) in comparison with each other observer separately was computed for each category. Thus 54 records per observer were available (six films, three observations for each of three categories) for the computation of each observer's timing tendencies with each of the other observers.[1] Bias will be indicated only if a large proportion of the samples for a given observer show the *same* direction in timing when he is paired with each of the other observers. For any given pair, the tendencies must obviously balance, since each was the standard of comparison for the other, and, even if all the samples for one of a single pair show a similar direction, it is impossible, on that basis alone, to decide which of the pair exhibits the bias.

Taking as a crude measure of bias, the proportionate distribu-

[1] As stated in Chapter VII, "timing tendency, as we define it, is synonymous with skewness when the norm is synchronous recording. We assume, in other words, that a given pair of observers will record most of the items of behavior in corresponding intervals, but that each of them will misplace some items by recording either in the interval preceding or in the interval following that in which the other observer records. We further assume that, in the absence of persistent idiosyncrasies, those items preceding the standard observer's record will tend to balance those following the standard observer's record, and that the timing tendency of each observer, as compared with other observers, can be indicated by his persistent misplacement of items in a particular direction."

tion of records which show retarded and accelerated tendencies, for each of the four observers, we find a high degree of individual consistency. The following table shows the number of records with retarded tendency, the number with accelerated tendency, and the number with balanced timing for each observer paired with each of the other three observers for three observations of each of the six films of Series A. Films, categories and observations are not differentiated in this summary.

TABLE XIX

Distribution of 54 Records of Observers W, X, Y and Z by Direction of Timing Tendency per Observer in all Pairings with other Observers (Series A)

Standard Observer	Co-Observers	Number of Records Showing: Retarded Timing	Accelerated Timing	Balanced Timing	Total Number of Records
W	X	14	39	1	54
	Y	18	34	2	54
	Z	53	1	0	54
	Total	85	74	3	162
X	W	39	14	1	54
	Y	29	19	6	54
	Z	52	2	0	54
	Total	120	35	7	162
Y	W	34	18	2	54
	X	19	29	6	54
	Z	53	0	1	54
	Total	106	47	9	162
Z	W	1	53	0	54
	X	2	52	0	54
	Y	0	53	1	54
	Total	3	158	1	162

Looking first at the totals per observer, we find that W shows no clear evidence of bias, the number of his accelerated records being about the same as the number of his retarded records. X has three times as many retarded as accelerated, Y more than twice as many retarded as accelerated, and Z no less than 158 accelerated records out of a total of 162. The three latter observers, then, show rather clear indications of bias, the extremes being Z (accelerated) and X (retarded).

Looking at the individual pairings, we find that Z is completely consistent with regard to all observers. X is also consistent in having appreciably more retarded than accelerated records with

each of the other observers, predominantly so when compared with Z, less so with W and Y. Y shows less consistency, being predominantly retarded in comparison with Z, twice as often retarded as accelerated with W, and actually more often accelerated than retarded when compared with X. W's apparent balance in the totals is made up of a predominant retardation when compared with Z, and marked acceleration when compared with both X and Y. The most consistently biased observers, then, from the standpoint of separate pairings, as well as of the totals, are Z and X.

We want to know also whether this consistency in direction of bias is accompanied by a similar consistency in *amount* of timing tendency. This can be indicated, in a very summary fashion, by examining the average amount of tendency by category and pair of observers for all films and observations combined and the average tendency for each observer in all possible pairings for the six films considered separately.

The following table gives the data for the first of these two comparisons, i. e., the average timing tendency per category for each pair of observers for all films and observations combined in Series A.

Standard Observer	Co-Observer	Material	Self	Person
W	X	−1.1	−0.8	−1.2
	Y	−1.7	−0.4	−1.1
	Z	+5.7	+2.6	+4.3
X	W	+1.1	+0.8	+1.2
	Y	+0.4	+0.3	0
	Z	+6.5	+3.1	+5.3
Y	W	+1.7	+0.4	+1.1
	X	−0.4	−0.3	0
	Z	+7.6	+2.8	+4.8
Z	W	−5.7	−2.6	−4.3
	X	−6.5	−3.1	−5.3
	Y	−7.6	−2.8	−4.8

In all three categories, the highest degrees of retardation are found for W, X and Y, each in comparison with Z, and, conversely, of course, the highest degrees of acceleration are found between Z and each of the other three observers. W shows a somewhat lesser degree of retardation in the Z combination than do the other two observers, in all three categories. The pairings

in which Z appears consistently show four to five times as great a timing tendency as do any of the other pairings.

The table on the next page shows each observer's average tendency with all other observers by separate films (the three observations for each film averaged).

The most noteworthy point about this table is the fact that so often Z's acceleration balances the sum of the retardation of the other three observers. This is true for Films I, II and VI in the *Material* category, for all six films in the *Self* category, and for Films I, II and IV in the *Person* category. Z is always decidedly the most accelerated observer, even when, as in Films III, IV and V, *Material*, and in Films III, V and VI, *Person*, some other observer is also accelerated.

X is the most retarded observer in Films III and IV and ties with Y for the highest place in Film I in the *Material* category. He is also highest in Films I, IV, V and VI in the *Self* category, and in Films I and V, in the *Person* category; whereas Y ranks highest in Films II, V and VI in *Material*. in Films II and III, *Self* and in Films II, IV and VI, *Person*. Both of these observers, then, appear to be highly retarded; each of them has but one accelerated film-category. W is only moderately retarded and actually shows accelerated average tendencies in 4 out of 18 film-categories.

The results of these several analyses seem to indicate that Z, at one extreme, is biased with a timing tendency toward acceleration, and X, at the other extreme, is biased with a tendency toward retardation; that Y is definitely retarded; and that W is relatively unbiased in the degree and direction of his timing tendency.

Percentages of timing error show the same indications of bias that we have noted in timing tendencies. The table on page 199 summarizes the average percentages of retarded and accelerated units for each observer in pairwise comparison with the other three observers, in the third observation of Series A. Observer Z has a predominant proportion of accelerated units for all films and categories; Observer X, a predominance of retarded units, except in Film III, *Person;* Observer Y, a predominance of retarded units, except in Film III, *Material;* Observer W has a larger per cent of retarded units, for all films and categories, with

Observer	Material						Self						Person					
	Films						Films						Films					
	I	II	III	IV	V	VI	I	II	III	IV	V	VI	I	II	III	IV	V	VI
W	+2.8	+1.4	+2.1	−0.6	−1.3	+2.1	+1.1	+0.3	+0.7	+0.2	+0.2	+0.5	+0.1	+1.6	+1.8	+0.9	−0.4	−0.5
X	+4.6	+0.2	+8.4	+5.4	+2.3	+0.5	+2.1	+1.0	+0.8	+1.2	+2.4	+1.1	+4.8	+1.5	−0.8	+1.7	+4.5	+2.1
Y	+4.6	+3.8	−0.4	+3.9	+2.8	+2.4	+0.8	+1.4	+0.9	+0.9	+1.0	+0.7	+2.7	+2.4	+0.5	+1.8	+1.2	+2.5
Z	−12.0	−5.4	−10.1	−8.7	−3.8	−5.0	−4.0	−2.7	−2.4	−2.3	−3.6	−2.3	−7.6	−5.5	−1.5	−4.4	−5.3	−4.1

Observer	Tendency	Material						Self						Person					
		Films						Films						Films					
		I	II	III	IV	V	VI	I	II	III	IV	V	VI	I	II	III	IV	V	VI
W	Ret.	7.7	8.3	9.3	6.0	6.6	8.9	5.0	3.6	3.4	3.1	4.4	3.9	6.0	4.5	3.8	3.5	5.2	4.0
	Acc.	5.9	6.4	6.0	6.7	6.7	6.6	3.6	3.2	2.4	2.6	4.1	3.2	5.9	3.1	1.6	2.3	5.6	4.3
X	Ret.	8.8	6.0	11.6	10.0	8.6	6.2	5.4	3.6	3.6	3.9	5.9	4.2	8.1	4.7	2.6	4.6	7.6	6.2
	Acc.	4.5	5.1	4.0	5.1	5.5	5.6	3.5	2.6	2.5	2.1	2.9	3.1	3.5	3.2	2.9	2.6	3.7	4.5
Y	Ret.	12.3	9.6	9.4	11.5	10.3	8.7	4.0	3.7	3.7	3.3	4.9	4.2	9.8	7.0	4.4	5.1	6.9	6.6
	Acc.	9.1	6.3	9.9	7.4	7.1	5.7	2.9	2.3	2.4	2.0	3.6	3.5	7.2	4.2	3.3	3.0	5.6	4.6
Z	Ret.	4.9	5.8	5.8	4.4	6.1	4.5	2.5	1.6	2.1	1.3	1.9	1.8	5.4	3.5	3.5	2.3	4.6	4.7
	Acc.	13.5	10.8	13.1	10.5	7.8	8.0	6.2	4.3	4.1	3.1	5.4	4.1	12.2	8.4	3.8	6.6	9.0	8.2

the exception of the *Material* category in Films IV and V, and the *Person* category in Films V and VI, but the difference between the per cents of retardation and of acceleration is much smaller than for the other observers. This observer approaches closest to the norm of an even balance between retarded and accelerated tendencies.

We conclude, from the foregoing discussion, that individual differences among observers in timing the same occurrences constitute an item of importance in the study of technical aspects of reliability of observation. In the investigation reported here the stage was definitely set for accurate timing, since the use of a single timepiece by all observers eliminated all timing discrepancy of a purely mechanical origin. Such disagreement as occurred, therefore, can be attributed directly to the chance errors or to the specific biases of the observer himself. The results of our analysis of direction of timing tendencies leave little doubt as to the fact that individual observers, judged by the relative standard of comparison with other observers, have clearly defined tendencies to err more in one direction than in the other, to retard or to accelerate predominantly their records of behavior. For lack of an adequate standard of the reality of the situations observed, we are unable to determine whether all four observers have distinct biases or whether the apparent excess tendencies of some are in reality merely the necessary consequence of the biases of others. The fact that Observer Z is accelerated in relation to all other observers and that Observer X is retarded with all observers, suggest that both are probably biased. If Z has an acceleration bias, we infer that the bias lies in his method of recording, since events cannot be seen earlier than they actually occur.

Fortunately, in view of the implications for observational recording in real life situations, biases toward retarding or accelerating the beginnings and endings of occurrences in continuous recording of the behavior of a given individual do not distort or falsify the general picture of the behavior, if they are regular and constant. A single observer recording in a life situation may have such a timing bias without its affecting the accuracy of his total record, if he times accurately the duration of occurrences within each category, and if the effect upon all observations is

equally distributed.[1] We have no proof, however, that the biases of the four observers are wholly constant.

Individual Patterns in Interpretation Shown in Simultaneous Observations of Different Observers

Unreliability has been measured by a coefficient derived from the number of time units in which an observer disagreed with other observers. Our concern, in the study of bias, is not with the *amount* of such disagreement but with its *pattern*. If observers disagree on a given number of units between any two categories, will each observer show an equal tendency to misinterpret (from the other observer's point of view) or will he show a definite tendency to favor one of the two categories at the expense of the other? We have already had evidence that observers vary in the proportions of behavior they record in the several categories.

An excess of a given category can be distributed in three possible ways:

(1) It may be partly a timing phenomenon, a tendency to extend the record in a given category after the character has left the screen. In such a case, there will be no disagreement with other observers, who will have ceased recording.

(2) It may be an inconsistent tendency, manifesting itself sometimes at the expense of one category, sometimes at the expense of the other, varying from one observation to another of the same situation, and from one situation to another.

(3) It may be a consistent tendency, manifesting itself in many situations and in consecutive observations of the same situation favoring one category at the expense of one or both of the others.

The two latter situations are the ones which interest us from the point of view of determining the existence of bias as a factor in the reliability of observers.

Since each observer has seen each of six films three times with three other observers, he may be paired by film, observation and co-observer in 54 combinations. We can get a crude indication

[1] In non-continuous recording of such intermittently recurring events as language, for example, there is a greater likelihood that timing bias will affect the total number of intervals of behavior since an event occurring on the border line between two intervals may be recorded as occurring in both, thus increasing the total.

of his tendency toward biased errors by finding the number of such combinations in which his disagreements favor one category at the expense of either or both of the others. The following table summarizes the distribution of the 54 observation-combinations per observer according to disagreements on the *Material-Self*, *Material-Person* and *Person-Self* categories.

Observer	Disagreements on the Material-Self Categories			Disagreements on the Material-Person Categories			Disagreements on the Person-Self Categories		
	No excess greater than 2 units	Excess of more than 2 units favoring		No excess greater than 2 units	Excess of more than 2 units favoring		No excess greater than 2 units	Excess of more than 2 units favoring	
		Material	Self		Material	Person		Person	Self
W	21	18	15	32	17	5	15	8	31
X	15	7	32	31	13	10	17	11	26
Y	15	23	16	24	10	20	16	24	14
Z	13	28	13	27	11	16	10	36	8

In their disagreements between the *Material* and *Self* categories, there were approximately balanced disagreements, i. e., neither category was favored by more than two units, in 21 out of 54 observation-combinations for W, 15 for X, 15 for Y and 13 for Z. Of the remaining observation-combinations, W had more than two units excess of *Material* in 18 and more than two units excess of *Self* in 15 observations; X had *Material* excesses in only 7, but *Self* in 32; Y showed *Material* excesses in 23 observation-combinations against *Self* excesses in 16; and Z's excesses were for *Material* and *Self*, respectively, 28 and 13. It might be inferred, therefore, that Z favors *Material* at the expense of *Self*, and X *Self* at the expense of *Material*, and that the other two observers show no evidence of biased errors between these two categories. Does Z also favor *Material* at the expense of *Person* and does X favor *Self* at the expense of *Person?* Z is almost exactly balanced on the excesses of *Material* and *Person;* in 27 observations the excesses either way are less than two units, in 11 there is a greater *Material* excess and in 16 a greater *Person* excess. We may therefore infer that he has no consistent *Material* bias and we look further at his *Person-Self* disagreements for evidence of a *Person* bias. Here we find only 10 observation-combinations where there is a balance of disagreements within two units

and 36 where *Person* is in excess against only 8 where *Self* is in excess. We may therefore conclude that Z shows a tendency to favor both *Material* and *Person* at the expense of *Self*, with possibly a stronger favoring of *Person* than of *Material*.

Returning to X, who was found to favor *Self* at the expense of *Material*, we find that he also tends to favor *Self* at the expense of *Person*, balanced disagreements (within two units) occurring in only 17 observation-combinations and *Person* excesses occurring in only 11 against 26 for *Self*. Looking now at his *Person-Material* disagreements, we find an almost exact balance, 31 observation-combinations showing no excess greater than two units for either category, against 10 *Person* excesses and 13 *Material*. X, therefore, may be said to show evidence of a definite favoring of *Self* at the expense of both other categories. Of the other two observers, W favors both *Material* and *Self* at the expense of *Person*, and shows a balance between *Material* and *Self* disagreements, and Y favors *Person* somewhat at the expense both of *Material* and *Self*, although his excesses are not as strongly indicative of bias in favor of *Person* (20 *Person* against 10 *Material;* 24 *Person* against 14 *Self*) as W's are indicative of bias against *Person* (17 *Material* against 5 *Person;* 31 *Self* against 8 *Person*).

It is, of course, possible that, with such a relatively small amount of data, chance plays an unduly large part in this sort of analysis, and that the overemphasis or underemphasis, even though it is found to occur fairly consistently in these observation-combinations, is more apparent than real. We can check our inferences further by examining the extent or proportion of such overemphasis or underemphasis but, because of the small number of disagreeing units involved, such a comparison cannot be made except for certain combinations of observations and observer-pairings. The table on the next page indicates the number of disagreeing units of each observer on each pair of categories for the total of all observation-combinations and for the total of the three observations of each of the six separate films, with the per cent favoring each category.

Summarizing, first, the results of the preceding analysis of the mere direction of excess for each observer for the 54 observation-combinations on each pair of categories: we found evidence that

TABLE XX

Number of Units of Disagreement on Paired Categories and Per Cent of Disagreeing Units Favoring Each of the Paired Categories for Each Film of Series A (three observations per film combined and all pairings for each observer combined) and for Total of All Films

Disagreements on the MATERIAL-SELF Categories

Observer	Number of Units of Disagreement — Films I	II	III	IV	V	VI	Total	Per Cent of Disagreeing Units Favoring Material — Films I	II	III	IV	V	VI	Total	Per Cent Favoring Self — Films I	II	III	IV	V	VI	Total
W	117	159	50	114	122	240	802	29.1	61.6	(−)	39.5	50.0	64.2	51.9	70.9	38.4	(+)	60.5	50.0	35.8	48.1
X	124	181	58	125	125	227	840	24.2	22.1	(−)	29.6	42.4	52.9	36.1	75.8	77.9	(+)	70.4	57.6	47.1	63.9
Y	156	164	46	121	168	256	911	87.2	65.2	(+)	65.3	40.5	30.5	54.3	12.8	34.8	(−)	34.7	59.5	69.5	45.7
Z	141	160	58	103	145	223	830	48.9	54.4	(+)	68.9	67.6	54.3	57.6	51.1	45.6	(−)	31.1	32.4	45.7	42.4

Disagreements on the MATERIAL-PERSON Categories

Observer	Number of Units of Disagreement — Films I	II	III	IV	V	VI	Total	Per Cent of Disagreeing Units Favoring Material — Films I	II	III	IV	V	VI	Total	Per Cent Favoring Person — Films I	II	III	IV	V	VI	Total
W	36	61	6	24	52	95	274	(+)	(+)	(−)	(+)	(+)	36.8	61.3	(−)	(−)	(+)	(−)	(−)	63.2	38.7
X	40	56	5	26	47	102	276	(+)	(−)	(+)	(−)	(+)	75.5	59.8	(−)	(+)	(−)	(+)	(−)	24.5	40.2
Y	36	58	8	33	73	105	313	(+)	(+)	(+)	(−)	11.0	41.0	38.0	(−)	(−)	(−)	(+)	89.0	59.0	62.0
Z	38	73	5	27	62	82	287	(−)	30.1	(−)	(−)	(+)	45.1	42.9	(+)	69.9	(+)	(+)	(−)	54.9	57.1

Disagreements on the PERSON-SELF Categories

Observer	Number of Units of Disagreement — Films I	II	III	IV	V	VI	Total	Per Cent of Disagreeing Units Favoring Person — Films I	II	III	IV	V	VI	Total	Per Cent Favoring Self — Films I	II	III	IV	V	VI	Total
W	141	292	77	157	99	303	1069	33.3	40.8	48.1	37.6	31.3	39.3	38.5	66.7	59.2	51.9	62.4	68.7	60.7	61.5
X	158	314	49	164	104	320	1109	24.7	27.4	(+)	52.4	41.3	51.9	40.1	75.3	72.6	(−)	47.6	58.7	48.1	59.9
Y	173	288	60	156	104	366	1147	71.1	57.6	(−)	60.9	59.6	41.5	54.1	28.9	42.4	(+)	39.1	40.4	58.5	45.9
Z	182	338	54	199	125	365	1263	64.8	72.5	(+)	49.2	64.0	65.8	64.6	35.2	27.5	(−)	50.8	36.0	34.2	35.4

Note:—Where the number of disagreeing units was considered too small for meaningful percentages, merely the direction of the excess is indicated. (+) = favoring, (—) against given category.

W favored both *Material* and *Self* at the expense of *Person*, that X favored *Self* against both *Material* and *Person*, that Y favored *Person* against both *Material* and *Self* and that Z favored both *Person* and *Material* at the expense of *Self*, with some indication that his bias was more strongly in favor of *Person* than *Material*.

Looking now at the proportions for each observer for the total of all observation-combinations and for the total of all three observations for all pairings for each of the six films considered separately: W's bias against *Person* is found to be proportionately significant[1] for his totals and to exist consistently in the six separate films. No less than 61.3% of his total disagreements on *Material-Person* favor *Material* and 61.5% of his total disagreements on *Person-Self* favor *Self*. The direction of *Material-Person* disagreements was consistent for all of the films except III, where there were only six units of disagreement altogether. The proportion of *Self*-favoring disagreements on the *Self-Person* categories was consistent for five of the six films, the range of percentages being from 59.2% to 68.7%. Film III, however, again deviated, showing approximately balanced disagreements. As between *Material* and *Self*, there is an approximate balance on the totals and great variation from strong favoring of *Self* (70.9%) on Film I, to strong favoring of *Material* (64.2% and 61.6%) on Films II and VI. This evidence confirms the previous inference that W is biased *against Person* and shows no real positive favoring of either *Material* or *Self*.

X's inferred *Self* bias also receives some confirmation. His proportion on *Material-Self* totals shows a 63.9% favoring of *Self;* on *Person-Self*, the proportion is a slightly lower (59.9%) favoring of *Self*. Five out of the six films confirm the *Self-Material* proportion found on the totals, the *Self*-favoring percentages ranging from 57.6% to 77.9% (three are over 70%), but Film VI shows a very slight favoring of *Material* (52.9%). The *Self-Person* situation is, however, not consistent for the individual films, only three out of six showing *Self*-favoring proportions. X tends to favor *Material* against *Person*, but again not consistently, the 59.8% on the total being confirmed in direction in

[1] By "proportionately significant" we mean that the percentage of units favoring one of two categories shall be significantly greater than the 50% which would be expected if the disagreeing units were equally distributed between the two categories.

only four of the six films and the small number of disagreements here make all proportions except the total probably indeterminate.

Y's *Person*-favoring proportions are shown by 62.0% *Person* on the *Person-Material* totals but only 54.1% on the *Person-Self* totals. There is considerable inconsistency with regard to the individual films in both comparisons. An apparent slight favoring of *Material* on the *Material-Self* disagreements (54.3%) is also unconfirmed when the separate films are examined. We may conclude, therefore, that Y shows no clear-cut tendency to biased disagreements.

Z's favoring of both *Material* and *Person* against *Self* is confirmed by rather strong and consistent proportions both in the totals and in the individual films. The totals show 57.6% *Material* in the *Material-Self* disagreements and 64.6% *Person* in the *Person-Self* disagreements. In both cases, five out of six films confirm the bias; the range for the five in the former is 54.3% to 68.9% *Material*, with Film I showing an approximate balance; in the latter, 64.0% to 72.5% for four films, whereas Film III agreed in direction but had too few disagreeing units to warrant a percentage and Film IV showed balanced disagreement. With regard to Z's favoring of *Person* against *Material*, this is also confirmed by a proportion of 57.1% on the totals and five out of six films agreeing on the direction of the excess. We may therefore conclude that Z's disagreements are biased both for *Person* and against *Self*.

This analysis, then, tends to confirm the previous inference of bias in both W and Z, partially to confirm X's bias and to throw doubt on the existence of any real bias on Y's part.

Two further points are of interest here: (1) Do the inferred biases manifest themselves equally strongly in all observer-pairings? (2) Does bias tend to diminish, persist, or increase with successive observations of the same situation? The table on the opposite page throws some light on both points.

Because of the small number of cases, it was impossible to make the pairwise comparison separately by films. For some of the pairs, however, there were enough disagreements to permit computations for three sub-samples, all first observations, all second observations and all third observations, respectively.

TABLE XXI

Number of Units of Disagreement on Paired Categories and Per Cent of Disagreeing Units Favoring Each of the Paired Categories, in First, Second and Third Observations of Series A (six films combined) for Separate Pairings of Each Observer and for All Pairings Combined

Disagreements on the Material-Self Categories

Observer	Co-observers	Number of Units of Disagreement: Observations 1	2	3	Total	Per Cent Favoring Material: Observations 1	2	3	Total	Per Cent Favoring Self: Observations 1	2	3	Total
W	X	116	79	55	250	64.7	70.9	(+)	64.4	35.3	29.1	(−)	35.6
	Y	123	114	52	289	45.5	56.1	(−)	48.8	54.5	43.9	(+)	51.2
	Z	117	82	64	263	33.3	56.1	(−)	43.3	66.7	43.9	(+)	56.7
	XYZ	356	275	171	802	47.8	60.4	46.8	51.9	52.2	39.6	53.2	48.1
X	Y	134	96	92	322	37.3	39.6	37.0	37.9	62.7	60.4	63.0	62.1
	Z	118	77	73	268	33.1	27.3	43.8	34.3	66.9	72.7	56.2	65.7
	WYZ	368	252	220	840	35.3	32.5	41.4	36.1	64.7	67.5	58.6	63.9
Y	Z	131	93	76	300	46.6	50.5	51.3	49.0	53.4	49.5	48.7	51.0
	WXZ	388	303	220	911	54.6	51.2	58.2	54.3	45.4	48.8	41.8	45.7
Z	WXY	365	252	213	830	62.2	54.8	53.1	57.6	37.8	45.2	46.9	42.4

Disagreements on the MATERIAL-PERSON Categories

Observer	Co-observers	Number of Units of Disagreement: Observations 1	2	3	Total	Per Cent Favoring Material: Observations 1	2	3	Total	Per Cent Favoring Person: Observations 1	2	3	Total
W	X	43	21	24	88	(−)	(+)	(+)	46.6	(+)	(−)	(−)	53.4
	Y	42	27	26	95	(+)	(+)	(+)	69.5	(−)	(−)	(−)	30.5
	Z	38	38	15	91	(+)	(+)	(−)	67.0	(−)	(−)	(+)	33.0
	XYZ	123	86	65	274	57.7	73.3	(+)	61.3	42.3	26.7	(−)	38.7
X	Y	56	21	28	105	(+)	(+)	(−)	63.8	(−)	(−)	(+)	36.2
	Z	26	32	25	83	(+)	(+)	(−)	61.4	(−)	(−)	(+)	38.6
	WYZ	125	74	77	276	72.8	58.1	40.3	59.8	27.2	41.9	59.7	40.2
Y	Z	44	40	29	113	(−)	(+)	(+)	46.0	(+)	(−)	(−)	54.0
	WXZ	142	88	83	313	27.5	39.8	54.2	38.0	72.5	60.2	45.8	62.0
Z	WXY	108	110	69	287	44.4	34.5	(+)	42.9	55.6	65.5	(−)	57.1

Disagreements on the PERSON-SELF Categories

Observer	Co-observers	Number of Units of Disagreement: Observations 1	2	3	Total	Per Cent Favoring Person: Observations 1	2	3	Total	Per Cent Favoring Self: Observations 1	2	3	Total
W	X	101	111	82	294	49.5	46.8	47.6	48.0	50.5	53.2	52.4	52.0
	Y	122	129	102	353	36.1	42.6	35.3	38.2	63.9	57.4	64.7	61.8
	Z	180	130	112	422	30.6	25.4	42.9	32.2	69.4	74.6	57.1	67.8
	XYZ	403	370	296	1069	37.0	37.8	41.6	38.5	63.0	62.2	58.4	61.5
X	Y	144	143	97	384	38.2	44.8	37.1	40.4	61.8	55.2	62.9	59.6
	Z	171	143	117	431	27.5	29.4	41.0	31.8	72.5	70.6	59.0	68.2
	WYZ	416	397	296	1109	36.8	41.6	42.9	40.1	63.2	58.4	57.1	59.9
Y	Z	174	138	98	410	41.4	37.0	52.0	42.4	58.6	63.0	48.0	57.6
	WXZ	440	410	297	1147	54.3	49.8	59.9	54.1	45.7	50.2	40.1	45.9
Z	WXY	525	411	327	1263	66.9	69.3	55.0	64.6	33.1	30.7	45.0	35.4

Notes: 1. Where the number of disagreeing units was considered too small for meaningful percentages, merely the direction of the excess is indicated, (+) favoring, (—) against a given category.

2. To avoid printing all percentages twice, figures are given here only once for each pair of observers. The percentages for X-W, for example, are the converse of those for W-X; e. g., in *Material-Self* disagreement, X's favoring of *Self* is found under W's favoring of *Material* to be 64.7%, etc.

W's bias against *Person* in the *Person-Material* disagreements is found not to exist at all in the WX pairing and to be quite strong in both the WY and WZ pairings. The same lack of bias when paired with X and strong bias when paired with Y and Z is found in the *Person-Self* disagreements. This result is confirmed, by and large, in the sub-samples.

X's bias favoring *Self* is found equally strongly for all three observer-pairings on the *Material-Self* disagreements, but on *Person-Self* disagreements strongest with Z, still appreciably strong with Y, and entirely absent with W. These results, also, are confirmed in the sub-samples.

Y's apparent *Person* bias is found strong with W and X, less strong with Z, on the *Material-Person* disagreements, and is found to be appreciably strong with W and X, but entirely lacking with Z on the *Person-Self* disagreements. These results tend to be confirmed by the sub-samples except in the *Person-Self* disagreements with Z.

Z's *Person*-favoring bias is found to be strong with W and X on all totals and most sub-samples, but inconsistent or absent with respect to Y. His favoring of *Material* at the expense of *Self* is inconsistent in his pairing with W, strong with X and absent with Y.

It thus appears that the several tendencies toward bias may be consistent when one of our observers is paired with each of the other three observers, as in X's favoring of *Self* in *Material-Self* disagreements, but in the other cases the general bias shown for total pairings is found to be definitely non-existent for certain specific pairings. Thus, we may conclude that W and X are unbiased with each other in respect to *Person-Self* and *Person-Material* disagreements, and Y and Z are also unbiased with each other in respect to *Person-Self* and *Person-Material* disagreements.

With regard to the question as to whether bias diminishes, persists or increases with consecutive observations, the totals for the six films combined represent the only adequate body of data.

W's bias against *Person* is found strongest in the second observation and weakest in the third in *Material-Person* disagreements, about equally strong in the first and second and weakest in the third on *Person-Self* disagreements.

X's *Self* bias is strongest in the second and weakest in the third on *Material-Self* disagreements, and approximately the same for all observations on *Person-Self*.

Y is variable both in direction and extent on consecutive observations.

Z shows least bias on the third observation in the *Person-Self* disagreements, and a reversal of bias in the third observation of *Material-Person*.

Thus, individual differences seem to play a rôle in the relative strength of bias in consecutive observations: X's bias shows perhaps the greatest persistence; Z's shows the greatest tendency to diminish.

Individual Patterns in Interpretation Shown in Consecutive Observations by the Same Observer

Where an observer is recording the same series of events in several consecutive observations, are his disagreements with himself in a given category randomly distributed or does an overemphasis developing in one part of an observation tend to reproduce itself throughout the observation?

The three consecutive records on each of the six films in Series A were analyzed for possible light on this question. Each film was divided into quarters, on a temporal basis (i. e., for an eighty-minute film where recording has been continuous the record for the period 0 to 20 minutes would be the first quarter, from 20 to 40 minutes, the second quarter, etc.). Slight discrepancies due to variation in the length of the consecutive observations, were distributed proportionately among the three categories. The number of units of each category was computed for each quarter of each film for each of the three consecutive showings, and the average for each category was calculated for Observers W, X, Y and Z, separately by quarters.

Each observer's deviations from his own average for a given category in a given film were then computed.[1]

The quarterly arithmetic averages and quarterly deviations from these averages for each observer are presented in Table XXII by film, category and observation. The first impression given by this table is that of a considerable tendency toward a carry-over of bias, as evidenced by a large number of cases in which a + (positive bias or overemphasis of a given category) appears in all four quarters of an observation, or three + quarters against only one − quarter. Similarly, four quarters with − deviations, (negative bias or underemphasis) or three − quarters against one + are frequent. Disregarding, for the moment, the size of the deviations, let us examine this apparent persistence, or carry-over, of bias purely on the basis of the direction (i. e., signs) of the deviations. There are 216 observation-categories divided into quarters for the four observers, or 54 for each observer. On the basis of chance arrangement, we would expect four + quarters, 1 in 16 times (6¼%), three + and one −, 4 in 16 times (25%), two + and two − , 6 in 16 times (37½%), three − and one +, 4 in 16 times (25%) and four −, 1 in 16 times (6¼%). Relating this expectancy to the group and to each individual, by allocating our actual findings to the appropriate categories,[2] we get the results which are given in table on page 213.

[1] To illustrate concretely:

Film VI, Series A Observer Y

Category	Quarter	Observation 1	Observation 2	Observation 3	Average for three observations	Deviations from average of three observations: Observation 1	2	3
Material	1	21.3	36.8	48.8	35.6	−14.3	+1.2	+13.2
	2	23.4	35.8	46.3	35.2	−11.8	+0.6	+11.1
	3	18.6	26.7	26.4	23.9	− 5.3	+2.8	+ 2.5
	4	27.9	34.2	27.9	30.0	− 2.1	+4.2	− 2.1
Self	1	198.2	181.9	168.1	182.7	+15.5	−0.8	−14.6
	2	214.0	194.7	186.0	198.2	+15.8	−3.5	−12.2
	3	229.0	215.1	228.3	224.1	+ 4.9	−9.0	+ 4.2
	4	223.6	216.6	223.9	221.4	+ 2.2	−4.8	+ 2.5
Person	1	65.5	66.2	68.2	66.6	− 1.1	−0.4	+ 1.6
	2	47.7	54.5	52.8	51.7	− 4.0	+2.8	+ 1.1
	3	37.4	43.1	30.4	37.0	+ 0.4	+6.1	− 6.6
	4	33.5	34.2	33.4	33.7	− 0.2	+0.5	− 0.3

[2] Zero deviations are distributed equally between +s and −s.

TABLE XXII

*Average Number of Units of Behavior in Three Observations of Each Quarter * of Each Film in Series A and Deviations from these Averages for Each of the Three Observations, by Observer and Category*

Observer	Film	Quarter	Material				Self				Person			
			Average	Deviation from Average: Obs. 1	Obs. 2	Obs. 3	Average	Deviation from Average: Obs. 1	Obs. 2	Obs. 3	Average	Deviation from Average: Obs. 1	Obs. 2	Obs. 3
W	I	1	5.5	− 1.6	+ 2.7	− 1.1	123.9	+ 8.6	− 7.4	− 1.2	32.5	− 7.1	+ 4.8	+ 2.3
		2	27.1	+ 0.3	− 3.4	+ 3.1	124.4	+ 1.6	− 1.2	− 0.4	10.1	− 1.6	+ 4.1	− 2.5
		3	12.8	− 5.3	+ 3.1	+ 2.2	115.5	+ 6.0	− 1.6	− 4.4	33.6	− 0.8	− 1.5	+ 2.3
		4	9.5	− 1.5	− 0.8	+ 2.3	122.4	+ 4.7	− 0.2	− 4.5	30.1	− 3.2	+ 0.9	+ 2.3
	II	1	55.8	− 5.1	+ 1.7	+ 3.4	161.7	+ 4.5	+ 1.5	− 6.0	80.5	+ 0.6	− 3.3	+ 2.7
		2	—	—	—	—	222.0	− 2.0	+ 2.1	− 0.1	76.0	+ 2.0	− 2.1	+ 0.1
		3	33.8	+ 1.2	+ 1.0	− 2.2	209.0	+ 3.4	− 0.9	− 2.5	55.2	− 4.5	− 0.3	+ 4.8
		4	10.4	− 0.7	+ 0.6	+ 0.1	236.6	+ 1.0	+ 0.8	− 1.8	51.0	− 0.3	− 1.6	+ 1.9
	III	1	0.8	− 0.9	+ 0.7	+ 0.2	136.3	+ 2.4	− 2.3	− 0.1	44.8	− 1.6	+ 1.6	0
		2	5.2	− 1.0	− 0.3	+ 1.3	170.3	− 0.9	− 0.2	+ 1.1	6.4	+ 1.9	+ 0.5	− 2.4
		3	7.8	0	+ 0.2	− 0.2	153.7	+ 6.0	− 4.9	− 1.1	20.5	− 6.0	+ 4.6	+ 1.4
		4	18.5	+ 0.7	+ 0.7	− 1.4	149.3	− 6.3	+ 3.6	+ 2.7	14.0	+ 5.8	− 4.6	− 1.2
	IV	1	23.3	− 2.8	+ 3.1	− 0.3	255.8	+ 3.5	− 3.2	− 0.3	39.3	− 0.8	+ 0.6	+ 0.2
		2	13.8	− 0.8	− 0.3	+ 1.1	274.4	− 0.7	+ 0.6	+ 0.1	30.0	+ 1.5	− 0.1	− 1.4
		3	28.2	+ 0.8	− 0.3	− 0.5	233.3	− 6.5	+ 2.7	+ 3.8	56.8	+ 5.7	− 2.5	− 3.2
		4	5.5	− 1.5	+ 2.5	− 1.0	221.3	+ 1.0	− 4.1	+ 3.1	91.5	+ 0.5	+ 1.8	− 2.3
	V	1	29.1	+ 0.4	+ 2.4	− 2.8	173.2	+ 0.8	− 2.5	+ 1.7	13.8	− 1.3	+ 0.2	+ 1.1
		2	21.2	+ 2.3	− 3.4	+ 1.1	172.2	− 2.7	+ 2.9	− 0.2	22.6	+ 0.4	+ 0.4	− 0.8
		3	10.9	− 2.5	+ 1.1	+ 1.4	185.8	+ 0.1	+ 0.9	− 1.0	19.2	+ 2.2	− 1.8	− 0.4
		4	25.5	+ 1.5	− 1.3	− 0.2	172.0	− 1.5	− 0.9	+ 2.4	18.5	0	+ 2.2	− 2.2
	VI	1	44.9	−25.4	+25.0	+ 0.4	170.7	+20.4	−22.5	+ 2.1	69.4	+ 5.1	− 2.4	− 2.7
		2	33.1	− 4.0	− 1.2	+ 5.2	196.2	+ 5.6	+ 0.9	− 6.5	55.6	− 1.5	+ 0.3	+ 1.2
		3	28.9	− 3.4	+ 1.5	+ 1.9	228.5	+ 3.4	− 2.0	− 1.4	27.5	+ 0.4	+ 0.2	− 0.6
		4	38.4	+ 0.9	− 1.9	+ 1.0	213.3	+ 2.3	− 0.2	− 2.1	33.3	− 3.2	+ 2.1	+ 1.1
X	I	1	7.7	+ 0.8	− 1.7	+ 0.9	130.5	− 1.7	+ 1.1	+ 0.6	23.7	+ 0.9	+ 0.5	− 1.4
		2	25.8	− 2.3	+ 2.4	− 0.1	125.1	+ 2.2	− 2.9	+ 0.7	11.1	0	+ 0.6	− 0.6
		3	10.4	− 0.8	− 1.7	+ 2.5	118.2	+ 1.1	+ 0.3	− 1.4	33.5	− 0.3	+ 1.4	− 1.1
		4	11.4	− 0.4	− 0.1	+ 0.5	119.4	+ 0.2	+ 0.1	− 0.3	31.2	+ 0.2	0	− 0.2
	II	1	39.4	−11.2	+ 3.7	+ 7.5	180.1	+12.4	− 3.5	− 8.9	78.4	− 1.3	− 0.1	+ 1.4
		2	—	—	—	—	218.7	+ 7.1	− 4.6	− 2.5	79.3	− 7.1	+ 4.6	+ 2.5
		3	32.4	− 1.8	+ 1.0	+ 0.8	212.6	− 0.4	+ 1.6	− 1.2	52.9	+ 2.1	− 2.6	+ 0.5
		4	5.4	− 0.9	+ 0.2	+ 0.7	242.7	+ 1.9	− 0.1	− 1.8	49.9	− 1.0	− 0.1	+ 1.1
	III	1	1.5	+ 1.0	− 0.5	− 0.5	135.8	− 1.4	+ 0.2	+ 1.2	44.7	+ 0.5	+ 0.2	− 0.7
		2	3.5	− 1.0	− 1.0	+ 2.0	171.8	+ 1.7	− 3.4	+ 1.7	6.7	− 0.7	+ 4.4	− 3.7
		3	9.3	+ 1.5	− 0.7	− 0.8	156.2	− 3.4	+ 3.1	+ 0.3	16.4	+ 1.7	− 2.3	+ 0.6
		4	17.7	+ 0.2	+ 1.0	− 1.2	144.8	+ 0.7	− 4.8	+ 4.1	19.5	− 0.9	+ 3.8	− 2.9
	IV	1	30.5	− 0.4	+ 0.6	− 0.2	249.4	− 0.6	+ 0.7	− 0.1	38.4	+ 1.0	− 1.4	+ 0.4
		2	15.0	− 0.2	− 0.7	+ 0.9	270.6	+ 0.9	− 0.4	− 0.5	32.7	− 0.6	+ 1.0	− 0.4
		3	23.4	+ 0.3	+ 0.3	− 0.6	236.1	+ 3.8	+ 5.5	− 9.3	58.8	− 4.1	− 6.0	+10.1
		4	5.1	− 0.7	+ 0.3	+ 0.4	218.7	− 1.5	+ 0.8	+ 0.7	94.5	+ 2.2	− 1.1	− 1.1
	V	1	28.7	+ 2.2	− 2.2	0	172.3	− 2.4	+ 1.6	+ 0.8	15.1	+ 0.1	+ 0.4	− 0.5
		2	21.8	− 1.1	− 1.3	+ 2.4	173.7	+ 0.3	+ 1.2	− 1.5	20.6	+ 0.6	− 0.1	− 0.5
		3	10.4	+ 0.2	− 0.5	+ 0.3	185.0	+ 2.6	+ 1.4	− 4.0	20.6	− 2.9	− 1.1	+ 4.0
		4	25.0	− 4.2	+ 1.0	+ 3.2	173.8	+ 2.6	− 1.4	− 1.2	17.3	+ 1.4	+ 0.3	− 1.7
	VI	1	41.8	0	+ 1.3	− 1.3	172.0	+ 0.6	− 6.8	+ 6.2	71.1	− 0.5	+ 5.2	− 4.7
		2	35.2	− 2.4	+ 3.4	− 1.0	192.3	− 0.5	+ 1.1	− 0.6	57.5	+ 3.0	− 4.6	+ 1.6
		3	25.3	− 7.2	+ 0.7	+ 6.5	230.1	+ 8.6	− 2.2	− 6.4	29.6	− 1.3	+ 1.4	− 0.1
		4	38.7	− 4.9	− 0.5	+ 5.4	213.0	+ 3.5	+ 2.1	− 5.6	33.3	+ 1.5	− 1.6	+ 0.1

* Corrected for slight discrepancies in timing. — No behavior. 0 No deviation from average

TABLE XXII—*Continued*

*Average Number of Units of Behavior in Three Observations of Each Quarter * of Each Film in Series A and Deviations from these Averages for Each of the Three Observations, by Observer and Category*

Observer	Film	Quarter	Material				Self				Person			
			Average	Deviation from Average			Average	Deviation from Average			Average	Deviation from Average		
				Obs. 1	Obs. 2	Obs. 3		Obs. 1	Obs. 2	Obs. 3		Obs. 1	Obs. 2	Obs. 3
Y	I	1	11.9	+ 0.1	+ 1.8	− 1.9	110.0	− 2.0	− 1.9	+ 3.9	40.1	+ 1.9	+ 0.1	− 2.0
		2	38.8	+ 1.7	+ 0.4	− 2.1	113.8	− 1.8	− 0.2	+ 2.0	9.4	+ 0.2	− 0.3	+ 0.1
		3	15.9	− 1.4	− 0.3	+ 1.7	106.6	+ 1.4	− 3.0	+ 1.6	39.5	0	+ 3.3	− 3.3
		4	15.5	+ 5.5	− 2.9	− 2.6	110.6	− 3.1	+ 3.6	− 0.5	35.9	− 2.4	− 0.8	+ 3.2
	II	1	52.6	−10.1	+ 0.5	+ 9.6	160.5	+ 8.4	− 0.3	− 8.1	85.0	+ 1.7	− 0.2	− 1.5
		2	0.3	− 0.3	+ 0.6	− 0.3	209.8	− 2.5	+ 4.5	− 2.0	88.0	+ 2.7	− 5.2	+ 2.5
		3	37.5	+ 1.5	− 0.8	− 0.7	196.0	− 1.4	+ 6.9	− 5.5	64.5	− 0.1	− 6.0	+ 6.1
		4	11.8	− 0.3	+ 0.6	− 0.3	236.6	−11.6	+10.4	+ 1.2	49.6	+ 1.8	− 1.0	− 0.8
	III	1	1.3	+ 0.7	+ 0.1	− 0.8	138.1	− 4.9	+ 2.5	+ 2.4	42.6	+ 4.2	− 2.6	− 1.6
		2	8.0	− 0.9	0	+ 0.9	166.9	+ 4.6	− 3.9	− 0.7	7.1	− 3.7	+ 3.9	− 0.2
		3	9.1	− 1.6	+ 1.8	− 0.2	156.3	+ 1.1	− 0.5	− 0.6	16.1	+ 0.6	− 1.3	+ 0.7
		4	19.6	− 1.6	− 0.1	+ 1.7	141.6	+ 1.7	− 0.6	− 1.1	20.8	− 0.1	+ 0.6	− 0.5
	IV	1	31.1	− 0.8	+ 2.0	− 1.2	247.0	+ 1.3	− 2.6	+ 1.3	40.3	− 0.5	+ 0.8	− 0.3
		2	16.3	+ 1.6	+ 0.7	− 2.3	266.6	− 3.9	− 1.0	+ 4.9	35.3	+ 2.5	+ 0.2	− 2.7
		3	29.7	+ 1.6	− 0.2	− 1.4	225.0	− 7.9	+ 2.8	+ 5.1	63.5	+ 6.5	− 2.6	− 3.9
		4	5.3	+ 0.2	− 0.3	+ 0.1	217.6	+ 2.7	− 4.8	+ 2.1	95.3	− 2.6	+ 5.1	− 2.5
	V	1	25.6	− 0.8	− 0.4	+ 1.2	172.3	+ 0.3	− 1.4	+ 1.1	18.1	+ 0.6	+ 1.8	− 2.4
		2	24.9	+ 2.3	− 0.1	− 2.2	173.8	− 2.2	− 0.4	+ 2.6	17.5	− 0.3	+ 0.6	− 0.3
		3	9.8	− 0.7	− 0.2	+ 0.9	182.1	+ 1.7	− 0.5	− 1.2	24.2	− 0.8	+ 0.6	+ 0.2
		4	18.5	− 1.8	− 2.0	+ 3.8	171.4	− 1.0	+ 0.4	+ 0.6	26.1	+ 2.7	+ 1.6	− 4.3
	VI	1	35.6	−14.4	+ 1.2	+13.2	182.7	+15.4	− 0.8	−14.6	66.6	− 1.1	− 0.4	+ 1.5
		2	35.2	−11.7	+ 0.6	+11.1	198.2	+15.8	− 3.5	−12.3	51.7	− 3.9	+ 2.8	+ 1.1
		3	23.9	− 5.3	+ 2.8	+ 2.5	224.1	+ 4.9	− 9.0	+ 4.1	37.0	+ 0.4	+ 6.1	− 6.5
		4	30.0	− 2.1	+ 4.2	− 2.1	221.4	+ 2.3	− 4.8	+ 2.5	33.7	− 0.2	+ 0.5	− 0.3
Z	I	1	10.7	+ 2.5	+ 0.6	− 3.1	116.3	− 0.4	− 0.6	+ 1.0	35.0	− 2.1	+ 0.2	+ 1.9
		2	28.7	+ 2.6	− 0.9	− 1.7	119.5	− 2.1	− 2.8	+ 4.9	13.8	− 0.6	+ 3.7	− 3.1
		3	9.3	− 2.8	+ 1.0	+ 1.8	112.9	− 1.6	− 1.7	+ 3.3	39.8	+ 4.3	+ 0.8	− 5.1
		4	13.7	+ 0.5	− 0.9	+ 0.4	112.2	− 4.3	+ 1.3	+ 3.0	36.2	+ 3.8	− 0.4	− 3.4
	II	1	48.9	−12.3	+ 3.5	+ 8.8	155.4	+ 4.5	− 8.8	+ 4.3	93.7	+ 8.0	+ 5.2	−13.2
		2	0.3	− 0.3	+ 0.6	− 0.3	210.6	− 9.2	+ 0.4	+ 8.8	87.0	+ 9.5	− 1.1	− 8.4
		3	32.2	− 5.0	+ 2.9	+ 2.1	194.7	+ 6.3	− 4.1	− 2.2	71.0	− 1.5	+ 1.4	+ 0.1
		4	10.0	+ 1.0	0	− 1.0	231.5	− 2.7	− 3.5	+ 6.2	56.4	+ 1.7	+ 3.5	− 5.2
	III	1	0.6	− 0.6	− 0.7	+ 1.3	134.8	+ 0.8	− 0.8	0	46.5	− 0.1	+ 1.7	− 1.6
		2	6.4	− 0.4	− 1.8	+ 2.2	168.3	+ 5.2	− 5.0	− 0.2	7.3	− 4.8	+ 6.9	− 2.1
		3	8.8	+ 0.7	+ 1.3	− 2.0	153.9	+ 0.4	− 3.9	+ 3.5	19.3	− 1.1	+ 2.5	− 1.4
		4	18.4	0	− 0.1	+ 0.1	142.8	− 1.5	+ 2.8	− 1.3	20.8	+ 1.3	− 2.5	+ 1.2
	IV	1	31.6	− 0.1	− 2.0	+ 2.1	242.9	− 3.1	+ 2.2	+ 0.9	43.7	+ 3.1	− 0.3	− 2.8
		2	17.3	+ 2.1	− 4.0	+ 1.9	264.3	− 1.0	+ 0.8	+ 0.2	36.7	− 1.1	+ 3.1	− 2.0
		3	30.2	+ 2.2	− 4.1	+ 1.9	227.6	− 1.7	+ 3.7	− 2.0	60.4	− 0.7	+ 0.5	+ 0.2
		4	4.1	+ 0.5	− 0.5	0	222.2	− 2.7	+ 1.4	+ 1.3	92.1	+ 1.9	− 0.5	− 1.4
	V	1	35.6	− 2.4	− 1.3	+ 3.7	164.1	+ 1.5	+ 0.3	− 1.8	16.2	+ 0.9	+ 1.0	− 1.9
		2	26.5	− 1.6	− 0.4	+ 2.0	167.6	+ 1.1	+ 0.4	− 1.5	21.8	+ 0.5	0	− 0.5
		3	10.8	+ 2.2	+ 1.0	− 3.2	178.3	− 0.7	− 1.8	+ 2.5	26.9	− 1.4	+ 0.8	+ 0.6
		4	29.0	+ 0.9	+ 0.6	− 1.5	167.6	− 2.2	− 5.5	+ 7.7	19.3	+ 1.3	+ 4.8	− 6.1
	VI	1	42.6	− 2.0	+ 3.5	− 1.5	159.9	− 2.3	+ 1.4	+ 0.9	82.5	+ 4.2	− 5.0	+ 0.8
		2	33.5	− 1.3	+ 2.2	− 0.9	187.2	+ 2.0	− 3.2	+ 1.2	64.2	− 0.8	+ 0.9	− 0.1
		3	31.1	−10.4	+ 5.9	+ 4.5	225.1	+11.2	− 8.0	− 3.2	28.8	− 0.8	+ 2.1	− 1.3
		4	38.6	+ 4.6	− 3.6	− 1.0	207.8	− 2.7	+ 3.1	− 0.4	38.5	− 2.0	+ 0.4	+ 1.6

* Corrected for slight discrepancies in timing. — No behavior. 0 No deviation from average

Group—216 Observation-Categories Observation-Categories with Quarters Having	Percentage Expected	Percentage Observed
4 +	6.25	7.5
3 + 1 −	25.00	26.0
2 + 2 −	37.50	32.0
3 − 1 +	25.00	27.0
4 −	6.25	7.5

Observers—54 Observation-Categories Each—Observation-Categories with Quarters Having	Percentage				
	Expected	Observed			
		W	X	Y	Z
4 +	6.25	8.0	7.0	7.0	6.0
3 + 1 −	25.00	25.0	28.0	24.0	27.0
2 + 2 −	37.50	38.0	29.0	28.0	33.0
3 − 1 +	25.00	20.0	30.0	32.0	28.0
4 −	6.25	9.0	6.0	9.0	6.0

For the group as a whole, there is a slight, but only possibly reliable, tendency toward persistence of bias, as indicated by the excess of the two classes of fours and the deficiency of the twos. The strong similarity of the observed to the expected percentages is, however, striking. For the individual observers, the extremes of 4 + and 4 − are of about the same order as the expected. Because of the small number of cases, it is impossible to attribute reliability to the apparent differences found for Observers W and Y.

It is worth while to look somewhat further at the data, however, to see if, within the classes considered, there is any individual patterning. For example, an individual observer might consistently compensate for an overemphasis in one quarter by an underemphasis in the following quarter. To what extent does this occur? The following table gives some slight evidence on the point:

N = 54 for each observer

Observer	Second quarter compared with first quarter		Third quarter compared with second quarter		Fourth quarter compared with third quarter	
	% same sign	% different sign	% same sign	% different sign	% same sign	% different sign
W	40.7	59.3	53.7	46.3	48.1	51.9
X	42.6	57.4	40.7	59.3	49.1	50.9
Y	60.2	39.8	53.7	46.3	45.4	54.6
Z	72.2	27.8	49.1	50.9	41.7	58.3

The only percentages in this table which can be considered significantly different from 50% are those for W, Y and Z when

the second quarter is compared with the first. W shows a possibly significant tendency to compensate in the second quarter for an overemphasis or underemphasis in the first quarter; Y shows a possibly significant, and Z, a rather definite, tendency to carry over an underemphasis or overemphasis from the first to the second quarters. The only other possibly significant percentages represent X's tendency to compensate in both the third and second quarters, and Z's final compensation in the fourth quarter. These differences are not worthy of very serious consideration because of the small number of cases used in computing the percentages. We may safely say, however, that so far as the signs of the consecutive quarterly deviations are a valid index of the persistence of bias over a whole observation, there is no marked tendency for bias to persist. There is some evidence that an overemphasis or underemphasis beginning in the first quarter tends to reproduce itself strongly in the second quarter for two of the observers, and for such first quarter overemphasis to be immediately compensated for by the other two observers.

It is probable that small deviations are quite insignificant on account of errors inherent in our original recording and the arbitrary nature of our later transcription.[1] It is possible that the only deviations of significance for this discussion of temporal bias are the relatively large deviations. If a single quarter is greater or less than the average for that quarter by (let us say, arbitrarily) two units, it is possible that the observer is making an interpretation significantly different from usual. And if, for a given category, deviations of this order and of the same sign appear consecutively in the four quarters, it seems probable that the individual observer is consistently making a biased interpretation in favor of a given category, whatever the situation presented to him by the film. Out of a total of 216 ± quarterly deviations for each observer, there was the following distribution of large deviations: W had 48 + and 45 −; X, 34 + and 32 −; Y, 45 + and 48 −; and Z, 50+ and 49 −.

Our interest centers on whether these significant deviations, which occur relatively evenly in all quarters, are found associated in the same observation and the same category. Looking at the distribution of the large positive deviations (indicating over-

[1] See Appendix A.

emphasis of a category) we find that in those observation-categories where any large positive deviation occurs in any one of the four quarters, there is only one such deviation in 38.9% of W's observation-categories, 40.7% of X's, 37.0% of Y's and 27.8% of Z's.

This indicates a possibly significant tendency for the large positive deviations of three of the four observers to occur discretely and to be unassociated with another large positive deviation in the same observation. The fourth observer (Z) shows some slight tendency toward a carry-over. This suggests that, on the whole, large errors, as measured from the observer's own consecutive records, do not influence one another. They seem to represent variations in interpreting specific instances and do not carry over to whole situations.

The complete distribution of deviations greater than two units in size is as follows:

Observation-Categories Containing	W	X	Y	Z
One + (no −)	8	14	9	8
One − (no +)	10	10	8	9
More than one + (no −)	7	4	7	8
More than one − (no +)	7	4	8	8
At least one + and one −	17	10	14	15

If we consider only those cases where there is more than one large deviation we find an approximately equal tendency for deviations combined in any given observation to be of the same sign and to be of opposite signs. We cannot, therefore, infer that there is any evidence that a large positive error produces a large compensatory negative error in the same observation, nor does it seem to produce a persistence of overemphasis.

This point is, on the whole, rather tenuous. The material is, perhaps, too scanty for the analysis we have given it. Our only conclusion is a tentative negative one; there is no clear evidence of what we may call temporal bias in our observers, but there is some suggestion of individual differences in this respect.

Individual Patterns in Interpretation Shown in Normal Records Compared with "True" Records

Our standard of comparison in all our preceding analyses of reliability has been the agreement of trained and competent

observers. We have assumed that an event is likely to have occurred in direct proportion to the number of such observers reporting its occurrence. From the point of view either of occurrence or of interpretation, all four observers have been found to agree on some of the behavior acts they are observing; three of them will agree on other acts; two on still other acts; and a sizeable proportion of items will be recorded or assigned to a specific category by only one of the four observers. It is reasonable to suppose, therefore, that they have all neglected to record certain acts that have actually happened, and that they have all recorded certain acts that have not actually happened. As we have seen, the proportions of the group agreeing in assigning continuous behavior to one of three mutually exclusive categories vary considerably. We have analyzed some of the individual patterns as they appear in the observer's inability to record the same event in the same way on consecutive observations and in his inability to agree either on timing or interpretation with trained observers recording simultaneously. We must also state the problem in terms of a possibly incomplete grasp of reality, or a tendency to neglect certain aspects of an event and to overemphasize others. If we had a "true" record of the event, we could obviously compare individual frequencies with the "true" frequencies and arrive at an estimate of this type of error. We have, of course, no such "true" record, but if the records of the slow speed motion picture films, Series C, are considered as an approximation to the "true" frequencies of the three behavior categories we have been studying, then a comparison of the relative frequencies of each of the categories in the normal speed records of this series with those of the slow speed records should indicate whether there is any consistent tendency on the part of our observers to overemphasize or neglect a particular category.

As it affects this aspect of the problem, bias may be defined as a general tendency (reproduced in individual patterns) to overemphasize any particular category at the expense of one or both of the other categories when the standard of comparison is an approximation to the "true" record of the behavior under observation.

There were three ordinary observations in Series C for both

normal and slow speed on each of three characters (Philip, John and Billy) where the films were sufficiently synchronized to permit a comparison of the corresponding records of the two speeds. For some of our comparisons we do not give the data in terms of the three characters separately.

Our observers are not in complete agreement on the interpretation of all units of behavior, even on the slow speed film. The average percentages of disagreeing units by category for the four observers and for the group, for the three characters, on the slow speed observation were as follows:

Observer	Material	Self	Person
W	3.6%	5.6%	9.7%
X	4.7%	6.0%	10.1%
Y	6.5%	6.6%	9.3%
Z	5.7%	4.2%	15.8%
WXYZ	5.2%	5.6%	11.4%

It seemed best, therefore, not to complicate our conception of the "true" distribution of units of categories by getting a composite distribution of all slow speed records for all observers, against which each observer's normal speed record could be compared. The factor of unreliability was, therefore, disregarded here, and an approximation to a "true" record was computed separately for each observer by averaging his own slow speed records for each character.

Two sorts of distortion of reality will be expected: (1) a negative distortion, i. e., the crude observations will be lacking in certain details which appear in the refined observations and (2) a positive distortion of reality, i. e., the crude observations will include certain details which do not appear in the refined observations. If these distortions of reality are unbiased, there should be no preponderance of any given kind of detail in either of these distortions. If these distortions of reality are biased, we will find a preponderance of particular sorts of details, either negatively or positively or both. Stated in terms of this technique and these observations of normal speed and slow speed motion pictures, a negative distortion of reality will be manifested by the proportion of each category of behavior recorded in the slow

observation and uncorroborated in the normal. The distortion will be considered biased if there are marked differences in numbers and proportions of units uncorroborated as between categories. Similarly, a positive distortion of reality will be indicated by the proportion of each category of behavior recorded in the normal observations and uncorroborated in the slow, and the distortion will, again, be considered biased if there are marked differences between categories.

By making the units of the normal observations exactly one-fourth those of the slow observations, each unit of the one was directly comparable with a corresponding unit of the other.[1]

On the opposite page, we present a tabulation derived by setting up a composite of each observer's three slow observations against a composite of his three normal observations. This table is in terms of amount, not proportion. It indicates the extent to which any unit in which *Material* was recorded in any (one, two or three) of the slow observations was corroborated by any corresponding unit of *Material* in any (one, two or three) of the normal observations, and, similarly, for the other two categories. Because of the extreme smallness of the normal unit, we have allowed for a possible corroboration, if an uncorroborated unit of either series was adjacent to a unit of the same category in the other series. We have, however, kept this doubtfully corroborated group separate in the tabulation. The final column, representing the totally uncorroborated units, is the most significant as an indication of the extent of distortion of reality.

The extent of negative distortion of reality in this technique is probably best expressed by the average per cent of uncorroborated units, or 6.2%. When we look at the actual number of such units, by category, however, we see that *Self* has more than twice as many sucn units as either *Material* or *Person*. Expressed as percentages of the total slow record for each category, however, *Person* and *Self* show about the same lack of corroboration, both being significantly greater than *Material*. All of the observers show the greatest actual *number* of uncorroborated units in the *Self* category, but both X and Y have a greater *proportion* of the *Person* category uncorroborated. X has also a large proportion

[1] See p. 124, *supra*.

of the *Material* category uncorroborated. Y has almost perfect corroboration in the *Material* and Z in the *Person* category. Thus, although a general tendency is found, individual differences among observers still persist.

TABLE XXIII

Number of Whole Intervals on Slow Records, Series C (three observations combined), with Number and Per Cent of Such Intervals Which Have Corresponding, Adjacent and No Corresponding or Adjacent Quarter Intervals of the Same Category on Normal Records (three observations combined) Totaled for Characters Philip, John and Billy, by Category and Observer

Category	Observer	All Slow Whole Intervals		Normal Quarter Intervals: Corresponding to Slow		Normal Quarter Intervals: Adjacent to Slow		Normal Quarter Intervals: Corresponding and Adjacent to Slow		Slow with No Corresponding or Adjacent Normal	
		Number	% of All Slow	Number	% of All Slow	Number	% of All Slow	Number	% of All Slow	Number	% of All Slow
Material	W	301	100.0	278	92.4	17	5.6	295	98.0	6	2.0
	X	309	100.0	265	85.8	19	6.1	284	91.9	25	8.1
	Y	326	100.0	315	96.6	8	2.5	323	99.1	3	0.9
	Z	315	100.0	287	91.1	20	6.4	307	97.5	8	2.5
	WXYZ	1251	100.0	1145	91.5	64	5.1	1209	96.6	42	3.4
Self	W	309	100.0	254	82.2	39	12.6	293	94.8	16	5.2
	X	313	100.0	222	70.9	51	16.3	273	87.2	40	12.8
	Y	294	100.0	221	75.2	47	16.0	268	91.2	26	8.8
	Z	274	100.0	218	79.6	39	14.2	257	93.8	17	6.2
	WXYZ	1190	100.0	915	76.9	176	14.8	1091	91.7	99	8.3
Person	W	99	100.0	78	78.8	17	17.2	95	96.0	4	4.0
	X	99	100.0	71	71.7	13	13.1	84	84.8	15	15.2
	Y	101	100.0	70	69.3	15	14.9	85	84.2	16	15.8
	Z	107	100.0	95	88.8	11	10.3	106	99.1	1	0.9
	WXYZ	406	100.0	314	77.3	56	13.8	370	91.1	36	8.9
All Categories All Observers		2847	100.0	2374	83.4	296	10.4	2670	93.8	177	6.2

Before passing on to the comparable tabulation for positive distortion of reality, let us try to estimate the degree to which each category was substituted for the others in this distorting process. For this purpose, we examined every unit in which there was no corroboration either in corresponding or adjacent units and recorded the units of each of the other two categories actually found. The table on the next page indicates, on that basis, the probable substitutions in units of each category for each other category.

When *Material* is uncorroborated, *Self* is substituted for it about three times as often as is *Person;* when *Self* is uncorrobo-

TABLE XXIV

Number of Units in Slow (Average) Observations Uncorroborated in Normal (Average) Observations and Their Probable Substitutions in Normal Observations (Series C)

Uncorroborated Units in	Observer	Probable Substitutions* in Units of			
		Material	Self	Person	Total
Material	W		4.5	1.5	6.0
	X		18.5	6.5	25.0
	Y		2.5	0.5	3.0
	Z		6.0	2.0	8.0
	Total		31.5	10.5	42.0
Self	W	14.5		1.5	16.0
	X	34.5		5.5	40.0
	Y	22.7		3.3	26.0
	Z	10.8		6.2	17.0
	Total	82.5		16.5	99.0
Person	W	1.7	2.3		4.0
	X	6.2	8.8		15.0
	Y	10.2	5.8		16.0
	Z	0.5	0.5		1.0
	Total	18.6	17.4		36.0

* These "probable substitutions" were estimated as follows: the units of each of the two other categories appearing in the interval corresponding to, preceding and following the uncorroborated unit were totaled for the two categories, and the relative proportions applied to the actual number of uncorroborated units.

rated, *Material* is substituted for it about five times as often as *Person;* when *Person* is uncorroborated, *Material* and *Self* tend to be substituted to an equal degree. The individual observers' patterns conform fairly well to the general group pattern except that W, X and Y show the strongest tendencies to substitute *Material* for *Self,* and Z approaches a balance between *Material* and *Person* in such substitutions.

The following table indicates the actual degree of positive distortion. The normal units were taken as standard and the number of such units which were uncorroborated on the slow record was computed. The total extent of such distortion amounts to 4.1% of the units.[1] Here are found about twice as many such units on the *Material* as on the *Self* category and over twice as many on the *Person* as on the *Material.* In terms of the proportion to the total category, the excess of *Person* is much greater than that of the other categories, no less than 13.8% of the units of that category being uncorroborated. It would seem, then,

[1] The positive and negative distortions will not necessarily balance each other because of the variation in the number of doubtfully corroborated, i. e., adjacent, units in each series (i. e., a unit will be considered adjacent in the normal series if it is next to a corroborated slow unit, whereas such a unit would not appear in the slow tabulation.)

that our observers have a positive bias favoring the preferred categories. The individual patterns conform to the group tendency, with the following exceptions: W has very few uncorroborated *Material* units and X and Z have both such large propor-

TABLE XXV

Number of Quarter Intervals on Normal Records, Series C (three observations combined) with Number and Per Cent of Such Intervals Which Have Corresponding, Adjacent, and No Corresponding or Adjacent Whole Intervals of the Same Category on Slow Records (three observations combined) Totaled for Characters Philip, John and Billy, by Category and Observer

Category	Observer	All Normal Quarter Intervals		Slow Whole Intervals						Normal with No Corresponding or Adjacent Slow	
				Corresponding to Normal		Adjacent to Normal		Corresponding and Adjacent to Normal			
		Number	% of All Normal	Number	% of All Normal	Number	% of All Normal	Number	% of All Normal	Number	% of All Normal
Material	W	313	100.0	278	88.8	31	9.9	309	98.7	4	1.3
	X	300	100.0	265	88.4	25	8.3	290	96.7	10	3.3
	Y	344	100.0	315	91.6	20	5.8	335	97.4	9	2.6
	Z	324	100.0	287	88.6	28	8.6	315	97.2	9	2.8
	WXYZ	1281	100.0	1145	89.4	104	8.1	1249	97.5	32	2.5
Self	W	293	100.0	254	86.7	35	11.9	289	98.6	4	1.4
	X	253	100.0	222	87.7	27	10.7	249	98.4	4	1.6
	Y	247	100.0	221	89.5	22	8.9	243	98.4	4	1.6
	Z	262	100.0	218	83.2	39	14.9	257	98.1	5	1.9
	WXYZ	1055	100.0	915	86.7	123	11.7	1038	98.4	17	1.6
Person	W	121	100.0	78	64.5	31	25.6	109	90.1	12	9.9
	X	121	100.0	71	58.7	24	19.8	95	78.5	26	21.5
	Y	97	100.0	70	72.2	18	18.5	88	90.7	9	9.3
	Z	145	100.0	95	65.5	30	20.7	125	86.2	20	13.8
	WXYZ	484	100.0	314	64.9	103	21.3	417	86.2	67	13.8
All Categories All Observers		2820	100.0	2374	84.2	330	11.7	2704	95.9	116	4.1

tions of uncorroborated *Person* units that between them they account for two-thirds of the total of such units.

This analysis, it must be remembered, takes no account of the weights of such distortion: a unit of the standard record is considered corroborated if a unit of the same category appears in the same or an adjacent interval on any one of the three observations of the record under comparison. The analysis has indicated a negative distortion biased in favor of *Self*, i. e., more of the *Self* category is uncorroborated than of the other two, and a positive distortion biased strongly in favor of *Person* and less strongly in

favor of *Material*. We must now examine the data from the point of view of the relative weights of these distortions.

The table on the opposite page shows, separately for each observer and for each character observed, the number of units of each category recorded in each of the three normal speed observations compared with the units recorded in each observer's slow speed standard observation (i. e., the average of his three slow speed observations).

The tendencies noted in the preceding analysis are, on the whole, confirmed. Twenty-five out of 36 of the *Material* normal speed deviations from the slow averages are positive, indicating overemphasis of this category; 29 out of 36 of the *Self* deviations are negative, indicating underemphasis; and 25 out of 36 of the *Person* deviations are positive, again indicating overemphasis. The average overemphasis of *Material* is highest for Y, his three normal observations showing about 11 units each of *Material* excess, when compared with his slow standard and amounting to about 16% of his total record of that category. His deviations are consistently high for his observations of all three characters and excessively so for his observations of Billy, the percentage excesses of the three observations of this character ranging from 30 to 35%. Six of the total of 11 negative deviations on this category occur in Z's observations; his average tendency ranges from a balance to a rather definite negative tendency, contrary to the general group bias.

The most consistent and high tendency on *Self* is again found for Y, who averages from 20 to 25% less of this category on his normal speed observations than his slow standard would lead one to expect. The underemphasis is rather general, but much slighter, for all the other observers.

The *Person* overemphasis amounts, on the average, to less units than did the *Material*, but the percentage overemphasis is considerably greater for all observers except Y. Z shows an excessive overemphasis of this category in his normal observations, ranging, in his averages, from six to eight units, and representing a percentage excess up to about 50%. Y shows the least consistent tendency toward excess, and in his first normal observation, there is actually a deficiency of *Person* units. X, too, shows a great deal of inconsistency, but tends toward a positive excess.

TABLE XXVI

Deviations in Units and Per Cent of Each of Three Normal Speed Observations from Average of Three Slow Speed Observations by Observer, Character Observed and Category, Series C.

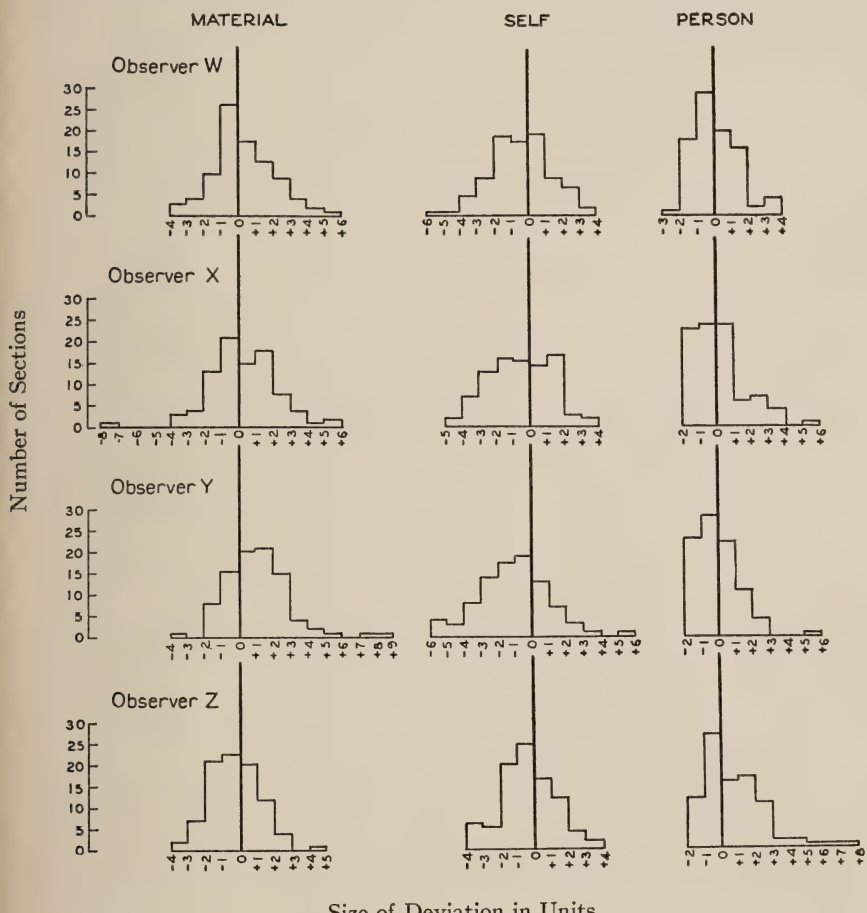
Category	Observer	Character Observed	Units Recorded in Normal Speed Observations			Units Recorded in Slow (Av. of 3 obs.)	Units of Deviation of Normal Speed Observations from Slow Average			Percentage Deviation of Normal Speed Observations from Slow Average		
			Obs. 1	Obs. 2	Obs. 3		Obs. 1	Obs. 2	Obs. 3	Obs. 1	Obs. 2	Obs. 3
Material	W	Philip	80.0	66.3	75.3	73.2	+ 6.8	− 6.9	+ 2.1	+ 9.3	− 9.5	+ 2.9
		John	69.0	65.0	67.5	61.7	+ 7.3	+ 3.3	+ 5.8	+11.8	+ 5.3	+ 9.4
		Billy	71.5	62.5	61.3	59.4	+12.1	+ 3.1	+ 1.9	+20.4	+ 5.2	+ 3.2
		Av.	73.5	64.6	68.0	64.8	+ 8.7	− 0.2	+ 3.2	+13.4	− 0.3	+ 4.9
	X	Philip	76.5	85.0	70.5	79.7	− 3.2	+ 5.3	− 9.2	− 4.0	+ 6.6	−11.6
		John	69.8	79.0	82.0	70.0	− 0.2	+ 9.0	+12.0	− 0.3	+12.9	+17.1
		Billy	59.5	48.0	64.3	49.8	+ 9.7	− 1.8	+14.5	+19.5	− 3.6	+29.1
		Av.	68.6	70.7	72.3	66.5	+ 2.1	+ 4.2	+ 5.8	+ 3.2	+ 6.3	+ 8.7
	Y	Philip	85.0	83.0	81.0	74.6	+10.4	+ 8.4	+ 6.4	+13.9	+11.3	+ 8.6
		John	76.0	74.0	75.5	70.2	+ 5.8	+ 3.8	+ 5.3	+ 8.3	+ 5.4	+ 7.5
		Billy	83.0	84.5	86.0	63.7	+19.3	+20.8	+22.3	+30.3	+32.7	+35.0
		Av.	81.3	80.5	80.8	69.5	+11.8	+11.0	+11.3	+17.0	+15.8	+16.3
	Z	Philip	75.0	79.3	75.0	81.0	− 6.0	− 1.7	− 6.0	− 7.4	− 2.1	− 7.4
		John	65.0	61.0	71.0	70.0	− 5.0	− 9.0	+ 1.0	− 7.1	−12.9	+ 1.4
		Billy	58.5	56.3	64.0	56.9	+ 1.6	− 0.6	+ 7.1	+ 2.8	− 1.1	+12.5
		Av.	66.2	65.5	70.0	69.3	− 3.1	− 3.8	+ 0.7	− 4.5	− 5.5	+ 1.0
Self	W	Philip	39.5	46.3	43.8	44.8	− 5.3	+ 1.5	− 1.0	−11.8	+ 3.3	− 2.2
		John	44.5	53.0	49.0	55.1	−10.6	− 2.1	− 6.1	−19.2	− 3.8	−11.1
		Billy	45.5	58.5	57.3	57.3	−11.8	+ 1.2	0.0	−20.6	+ 2.1	0.0
		Av.	43.2	52.6	50.0	52.4	− 9.2	+ 0.2	− 2.4	−17.6	+ 0.4	− 4.6
	X	Philip	36.5	29.0	41.0	37.7	− 1.2	− 8.7	+ 3.3	− 3.2	−23.1	+ 8.8
		John	44.3	42.5	40.0	48.5	− 4.2	− 6.0	− 8.5	− 8.7	−12.4	−17.5
		Billy	56.0	63.5	54.8	64.3	− 8.3	− 0.8	− 9.5	−12.9	− 1.2	−14.8
		Av.	45.6	45.0	45.3	50.2	− 4.6	− 5.2	− 4.9	− 9.2	−10.4	− 9.8
	Y	Philip	33.5	30.5	36.5	43.1	− 9.6	−12.6	− 6.6	−22.3	−29.2	−15.3
		John	43.0	41.5	38.5	47.1	− 4.1	− 5.6	− 8.6	− 8.7	−11.9	−18.3
		Billy	38.5	35.5	37.5	53.2	−14.7	−17.7	−15.7	−27.6	−33.3	−29.5
		Av.	38.3	35.8	37.5	47.8	− 9.5	−12.0	−10.3	−19.9	−25.1	−21.5
	Z	Philip	32.5	33.3	29.0	33.0	− 0.5	+ 0.3	− 4.0	− 1.5	+ 0.9	−12.1
		John	45.5	50.5	40.5	47.4	− 1.9	+ 3.1	− 6.9	− 4.0	+ 6.5	−14.6
		Billy	56.0	46.8	46.0	55.6	+ 0.4	− 8.8	− 9.6	+ 0.7	−15.8	−17.3
		Av.	44.7	43.5	38.5	45.3	− 0.6	− 1.8	− 6.8	− 1.3	− 4.0	−15.0
Person	W	Philip	13.5	17.3	14.3	12.2	+ 1.3	+ 5.1	+ 2.1	+10.7	+41.8	+17.2
		John	19.0	15.0	15.5	14.2	+ 4.8	+ 0.8	+ 1.3	+33.8	+ 5.6	+ 9.2
		Billy	16.0	12.0	14.3	14.3	+ 1.7	− 2.3	0.0	+11.9	−16.1	0.0
		Av.	16.2	14.8	14.7	13.6	+ 2.6	+ 1.2	+ 1.1	+19.1	+ 8.8	+ 8.1
	X	Philip	19.0	19.0	21.5	14.3	+ 4.7	+ 4.7	+ 7.2	+32.9	+32.9	+50.4
		John	18.8	11.5	11.0	13.0	+ 5.8	− 1.5	− 2.0	+44.6	−11.5	−15.4
		Billy	13.5	20.5	14.8	16.5	− 3.0	+ 4.0	− 1.7	−18.2	+24.2	−10.3
		Av.	17.1	17.0	15.8	14.6	+ 2.5	+ 2.4	+ 1.2	+17.1	+16.4	+ 8.2
	Y	Philip	13.5	17.5	15.5	13.2	+ 0.3	+ 4.3	+ 2.3	+ 2.3	+32.6	+17.4
		John	13.0	15.5	17.0	14.3	− 1.3	+ 1.2	+ 2.7	− 9.1	+ 8.4	+18.9
		Billy	12.5	12.0	10.5	13.1	− 0.6	− 1.1	− 2.6	− 4.6	− 8.4	−19.8
		Av.	13.0	15.0	14.3	13.5	− 0.5	+ 1.5	+ 0.8	− 3.7	+11.1	+ 5.9
	Z	Philip	25.5	18.3	28.0	15.2	+10.3	+ 3.1	+12.8	+67.8	+20.4	+84.2
		John	22.5	20.5	21.5	13.5	+ 9.0	+ 7.0	+ 8.0	+66.7	+51.9	+59.3
		Billy	16.5	29.8	23.0	18.8	− 2.3	+11.0	+ 4.2	−12.2	+58.5	+22.3
		Av.	21.5	22.9	24.2	15.8	+ 5.7	+ 7.1	+ 8.4	+36.1	+44.9	+53.2

There seems, then, to be a definite group tendency toward bias, with some individual differences in emphasis, when normal speed records are compared with approximations to "true" records. The direction of this bias is surprising, since it is the preferred categories (the most difficult) which are overestimated and the *Self* category which is underestimated. It will be well to examine the data further with regard to this tendency since the totals may prove unrepresentative of the successive situations or blocks of behavior found within the records. The average slow record for each observer was divided into consecutive tenths of each category. Each category and each observer were treated separately for each character observed. The boundaries of each tenth of *Material* were found for each of the three characters and observers, then the boundaries of each tenth of *Self* and, finally, the boundaries of each tenth of *Person*. These boundaries were marked off on each of the corresponding normal speed records and the number of units of the given category in each section was calculated. The number of such quarter intervals in each section of the normal speed film would have equaled one-tenth the number of whole intervals in the slow speed film if the normal recording had been adequate. If there were no bias, the deviations of the normal sections from the slow tenths would tend to be distributed normally. Our analysis of totals has shown that bias exists and our concern now is to see to what extent it manifests itself in specific situations, since each of these sections of a record represents a constellation of different behavior acts. We are interested to see whether the total bias is made up of a few large items or of many smaller items throughout the whole of the observation, whether it repeats itself in many sections or is found only in a few.

There were thirty sections (representing three characters observed) for each observer for each of three normal speed observations of the film. Histogram Series VII and the table following show the distribution of the quarter-interval deviations. The distributions are definitely skewed for the group, negatively for *Self* and positively for *Material* and *Person*. We can consider deviations of between 0 and $\pm$ 1 quarter interval as of little significance and partly due to our method of recording and transcription; deviations between 1 and 2 quarter intervals as

also partly reflecting our arbitrary recording; and those of 2 or more quarter intervals as possibly significant. The table on

Histogram Series VII. Distribution of Quarter-Interval Unit Deviations of Sections of Normal Speed Observations Corresponding to Actual Tenths of Slow Speed (Average) Observations, by Size of Deviation in Units, Observer and Category, Series C.



page 227 indicates the extent and direction of bias. Looking first at the total 360 deviations per category for the group (4 observers, 3 characters observed, 3 observations, by 10 sections each) we find 44.2% of the deviations in the *Material* category insig-

TABLE XXVII

Distribution of Quarter-Interval Unit Deviations of Sections of Normal Speed Observations Corresponding to Actual Tenths of Slow Speed (Average) Observations, by Size of Deviation in Units, by Observer and by Category (3 Characters, 3 Observations, Series C)

Unit Deviations	Material					Self					Person				
	Observers					Observers					Observers				
	W	X	Y	Z	Group	W	X	Y	Z	Group	W	X	Y	Z	Group
+8			1		1										
+7			1		1									1	1
+6														1	1
+5	1	2	1		4			1		1		1	1	1	3
+4	2	1	2	1	6									2	2
+3	4	4	4		12	2	2	1	2	7	4	4		2	10
+2	9	8	15	4	36	7	3	3	4	17	2	7	4	11	24
+1	13	18	21	12	64	9	17	7	12	45	16	6	11	17	50
*0 to +1	17½	15	20½	20½	73½	19½	14½	13	16½	63½	20	24½	22½	16	83
*0 to −1	26½	21	15½	22½	85½	17½	15½	19	24½	76½	29	24½	28½	27	109
−1	10	13	8	21	52	19	16	17	20	72	18	23	23	12	76
−2	4	4		7	15	9	13	14	5	41	1				1
−3	3	3	1	2	9	5	7	8	6	26					
−4						1	2	3		6					
−5						1		4		5					
−6															
−7		1			1										
−8															
All Deviations	90	90	90	90	360	90	90	90	90	360	90	90	90	90	360

*Zero deviations are divided equally between these two classes.

nificant, i. e., with no indication of the expected positive bias. The deviations of from + 1 to + 2 comprise 17.8% and those of − 1 to − 2, 14.4%, a scarcely significant balance favoring positive bias. The deviations of more than + 2, however, are 16.7% of the whole, compared with 6.9% for deviations of more than − 2, a significant balance in favor of the positive bias. It is obvious, however, that the bias shown in the totals does not manifest itself in all situations under observation but is made up of an overestimation of a relatively small portion of the separate events or sections of behavior under observation, and that an underestimation of this category in certain sections is not infrequent. The individual observers, for whom the base in all cases is 90, show the same general tendencies as the group, with the exception of Z, who has an excess of large negative deviations over large positive (10.0% against 5.6%). Y's favoring of *Material* is more general throughout his observations and he is again the most extreme of the *Material*-favoring observers, W, X and Y, i. e., he has 23,3% positive against 8.9% negative deviations of the order of 1 to 2 quarter intervals and 26.7% positive against 1.1% negative in the larger deviations.

TABLE XXVIII

Number and Per Cent of Total Quarter-Interval Unit Deviations of Sections of Normal Speed Observations Corresponding to Actual Tenths of Slow Speed (Average) Observations, by Degree of Bias, Category and Observers. (3 Characters, 3 Observations, Series C.)

Category	Material					Self					Person				
Degree of Bias and Range of Deviations for Each Degree	Observers					Observers					Observers				
	W	X	Y	Z	All	W	X	Y	Z	All	W	X	Y	Z	All
Number of Sections with															
No Bias															
(0 to ±1)	44	36	36	43	159	37	30	32	41	140	49	49	51	43	192
Possible Bias															
(+1 to +2)	13	18	21	12	64	9	17	7	12	45	16	6	11	17	50
(−1 to −2)	10	13	8	21	52	19	16	17	20	72	18	23	23	12	76
Probable Bias															
(+2 & over)	16	15	24	5	60	9	5	5	6	25	6	12	5	18	41
(−2 & over)	7	8	1	9	25	16	22	29	11	78	1	0	0	0	1
Total	90	90	90	90	360	90	90	90	90	360	90	90	90	90	360
% of Total Sections with															
No Bias															
(0 to ±1)	48.9	40.0	40.0	47.8	44.2	41.1	33.3	35.6	45.6	38.9	54.4	54.4	56.7	47.8	53.3
Possible Bias															
(+1 to +2)	14.4	20.0	23.3	13.3	17.8	10.0	18.9	7.7	13.3	12.5	17.8	6.7	12.1	18.9	13.9
(−1 to −2)	11.1	14.4	8.9	23.3	14.4	21.1	17.8	18.9	22.2	20.0	20.0	25.6	25.6	13.3	21.1
Probable Bias															
(+2 & over)	17.8	16.7	26.7	5.6	16.7	10.0	5.6	5.6	6.7	6.9	6.7	13.3	5.6	20.0	11.4
(−2 & over)	7.8	8.9	1.1	10.0	6.9	17.8	24.4	32.2	12.2	21.7	1.1	0.0	0.0	0.0	0.3
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

In the *Self* category, 38.9% of all deviations of all observers are insignificant, 20.0% are between − 1 and − 2 quarter intervals against 12.5% between + 1 and + 2, and 21.7% as great as or greater than − 2 against only 6.9% as great as or greater than + 2. This represents a significant tendency toward underemphasis. There is, then, a somewhat more prevalent tendency toward underemphasis of *Self*, than toward overemphasis of *Material*, as indicated by extensiveness throughout many events and sections of behavior. The individual observers are consistently in line with the group tendency with the exception of X, who shows an approximate balance of positive and negative deviations of the order of 1 to 2 quarter intervals, but an excess of negative over positive in the higher deviations. Y again shows the most definite tendency for the bias to manifest itself throughout the many behavior-sections of his observations.

In the *Person* category, more than half the deviations (53.3%) are insignificant for the group. The deviations of the order of 1 to 2 quarter intervals showed an excess of negative over positive, 21.1% against 13.9% of the total, but the deviations as great as or greater than 2 quarter intervals showed 11.4% positive against 0.3% negative. It is clear, then, that our assumption of a general bias in the *Person* category is incorrect; the apparent bias is due to the overestimation of a few sizeable items and under emphasis is actually more frequently found if portions of the sequence are considered. All the observers show similar tendencies, but Z has the most extreme and consistent overemphasis.

The excess of "large" deviations may be summarized as follows:

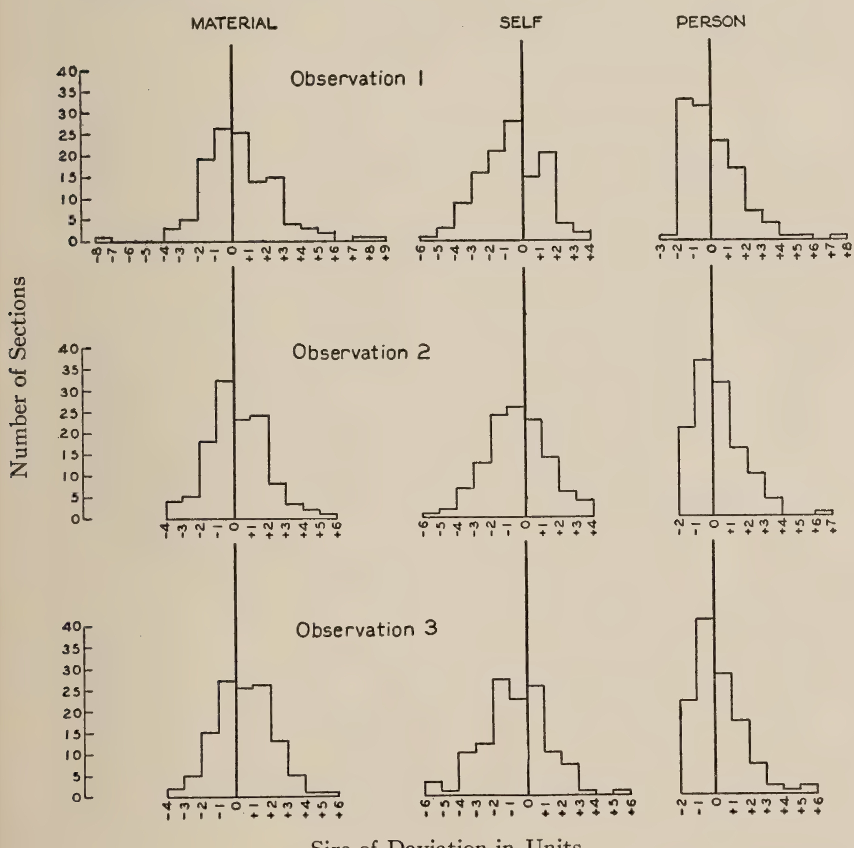
In the *Material* category, there were 60 deviations of + 2 or greater and 25 of − 2 or less. Of these, Z had 4 more negative than positive. All the others had excesses of positive over negative; Y had 23, W, 9 and X, 7.

In the *Self* category, there were 25 deviations of + 2 or greater, and 78 of − 2 or less. Of these, each observer contributed more negative than positive. Y had the largest excess of negative over positive, 24. X had 17, W, 7 and Z, 5.

In the *Person* category, there were 41 deviations of + 2 or greater and only one less than − 2. Of the excess of positive over negative deviation, Z contributed 18, X, 12, W, 5 and Y, 5.

One further point is of interest. Does the group tendency toward bias diminish with consecutive observations, or is it a

Histogram Series VIII. Distribution of Quarter-Interval Unit Deviations of Sections of Normal Speed Observations Corresponding to Actual Tenths of Slow Speed (Average) Observations, by Size of Deviation in Units, Observation and Category, Series C.



constant and consistent tendency manifesting itself to about the same degree when there are several consecutive observations of a series of situations? The observations in Series C are not entirely appropriate for answering such a question, as the mechanical difficulties were so great in getting the experiment

under way that a certain amount of unanticipated practice occurred in those observations which had to be discarded because of faults in the technique of projection, etc. With the reservation in mind that the so-called first, second and third observations were made after a practice period with the same film, we present the following table, where the units of deviation, by sections, of the normal observations from tenths of the slow observations, are given separately for the first, second and third observation. These data are plotted in Histogram Series VIII.

TABLE XXIX

Distribution of Quarter-Interval Unit Deviations of Sections of Normal Speed Observations Corresponding to Actual Tenths of Slow Speed (Average) Observations, by Size of Deviation in Units, by Observation and by Category (3 Characters, 4 Observers, Series C)

Unit Deviations	Material				Self				Person			
	Observations				Observations				Observations			
	First	Second	Third	All	First	Second	Third	All	First	Second	Third	All
+8	1			1								
+7	1			1					1			1
+6										1		1
+5	2	1	1	4			1	1	1		2	3
+4	3	2	1	6					1		1	2
+3	4	3	5	12	2	4	1	7	4	4	2	10
+2	15	8	13	36	4	6	7	17	7	10	7	24
+1	14	24	26	64	21	14	10	45	17	16	17	50
*0 to +1	25½	23	25	73½	15	23	25½	63½	23½	31½	28	83
*0 to −1	26½	32	27	85½	28	26	22½	76½	31½	36½	41	109
−1	19	18	15	52	21	24	27	72	33	21	22	76
−2	5	5	5	15	16	13	12	41	1			1
−3	3	4	2	9	9	7	10	26				
−4					3	2	1	6				
−5					1	1	3	5				
−6												
−7	1			1								
−8												
All Deviations	120	120	120	360	120	120	120	360	120	120	120	360

*Zero deviations are divided equally between the two classes.

The table gives a general picture of greater bias in the first than in succeeding observations for both *Material* and *Person*, but not for *Self*. When, however, the classes are examined in detail, the differences are found to be slight and probably unreliable. In the *Material* category, there are 26 sections with deviations as great as or greater than + 2 in the first observation, 14 in the second and 20 in the third; in the *Self* category, there are 29 as great as or greater than − 2 in the first, 23 in the second and 26 in the third observation; and in the *Person* category, there are

14 as great as or greater than + 2 in the first, 15 in the second and 12 in the third observation. Nor is there any very definite increase in the sections with clearly insignificant deviations, i. e., 0 to ± 1. The number of such sections in the *Material* category is 52, 55, 52 for first, second and third observations, respectively. The comparable numbers for *Self* are 43, 49 and 48, respectively, and for *Person*, 55, 68 and 69, respectively. Whether this means that bias with regard to reality does not tend to diminish with consecutive observations, or merely that the previous unexpected practice makes the comparison invalid, we are, unfortunately, not in a position to determine.

Summarizing briefly in terms of the questions asked in the beginning of the chapter:

(1) The timing tendencies of our four observers were found to repeat themselves in the many samples studied. Z showed a strong, and highly consistent, tendency toward acceleration, when compared with other observers, and X and Y quite consistent, but less strong, tendencies toward retardation.

(2) The interpretation disagreements of these observers formed rather clear-cut patterns, repeating themselves in many observational samples. W tended to favor both *Material* and *Self* at the expense of *Person;* X favored *Self*, though not with complete consistency, at the expense of both *Material* and *Person;* and Z favored *Person* at the expense of both other categories and also favored both *Person* and *Material* against *Self*.

(3) There was no real evidence that for any of our observers an overemphasis, in comparison with the observer's own established tendencies, developing early in any observation, tended to persist throughout the observation.

(4) When compared with approximations to "true" records, the "true" record always being defined individually for each of the observers, our whole group showed tendencies toward systematic distortions of reality, particularly in an overemphasis of *Person*, and an underemphasis of *Self*. Individual differences in patterns persisted here, too, especially in Y's strong overemphasis of *Material* (there was not complete agreement in the group but a general tendency also for W and X to overemphasize *Material*), and Z's excessive overemphasis of *Person*.

CHAPTER X

The Situation as a Factor in Observer Reliability

A complete analysis of the functioning of this observational technique would call for an estimate not only of the extent to which observers differ with one another on a series of observations, the extent to which they disagree with a standard, and the extent to which they show self consistency, but also an estimate of the extent to which the type of situation observed contributes to the observational error. We must attempt to answer, as best we can, the following questions.

To what extent do individual differences in the situations observed determine the degree of reliability of a group of observers, when reliability is measured by agreement among such observers? Or, stated somewhat differently, what is the degree of variability in the efficiency of the technique from one situation to another? If variability be found in the efficiency with which the technique and observers function, what factors in the situations contributing to the variability can be described, isolated and measured?

Our data are clearly not suitable for any thorough analysis of these points. The records made by four observers do not provide a statistically secure basis for determining the relative observational difficulty of the situations observed, and the rather definite individual tendencies of these observers further complicate the problem. It seems worth while, however, to go as far as we can with the data at hand, if for no other reason than to recognize situational variability as a factor that must be taken into account.

The best data for this purpose are the records of the third observation of each of the films in Series A. Since the observers showed rather consistently greater agreement on the third than on preceding observations, any tendencies toward real situational variability (as isolated from individual observers' variability) should be more apparent in this observation than in the less reliable ones.

In the preceding chapters on timing and interpretation, an

analysis was made of the differing reliability of the observations of the three behavior categories *Material*, *Self*, and *Person*. *Material* was found generally to be recorded with the least reliability, *Person* next most reliably and *Self* by all odds most reliably. It will be remembered that material behavior was differentiated from self and person behavior in such a way as to be given preference in marginal situations, i. e., if the character under observation was simultaneously manipulating material and in physical contact with another person, the behavior was, by definition, classified as *Material*, rather than as *Person;* if he was manipulating material and at the same time talking, walking about or engaged in other so-called self activities, the behavior was classified as *Material* rather than as *Self*. *Person* was also, by definition, given preference over *Self* in marginal situations. That these marginal situations were more difficult to classify than clear-cut discrete behavior acts seems unquestionable, but that they should operate disproportionately at the expense of the preferred category, other things being equal, would not seem, on *a priori* grounds to be necessary. That "other things" were not equal is clear from our preceding analyses. The amount of behavior in each of the preferred categories tends to be consistently much less than that in the non-preferred category, thus causing a differential effect of these marginal situations upon the preferred and the non-preferred category, due to the mere difference in the size of the base. Since we have, unfortunately, no measure of the frequency of marginal situations in the different films in our series, this factor cannot be isolated, but the nature of its effect can be inferred. In the persistent differences between categories, we found the amount of behavior in the category to be positively correlated with the reliability of the observers, a relationship which undoubtedly resulted from the operation of these marginal situations against the category with the small base. If these situations did not occur with the same relative frequency in the several films, there is no reason to suppose that variations from film to film in the amount of behavior recorded in a given category will necessarily show a correlation with reliability, in so far as reliability is the resultant of this factor. If such a correlation is found, we may infer the predominance and relative stability of this factor throughout all the films. If

no such correlation exists, we may infer that this factor is itself a variable whose contribution to situational difficulty cannot be measured indirectly by the variability in the amount or proportion of behavior of a given category.

Referring back to the interpretation chapter, we can relate this variability to the interpretation unreliability which is summarized in the following abstract from Table XVII.

Indices of Interpretive Disagreement by Observer, Film and Category for the Third Observation of Each Film in Series A

Category	Observer	Films					
		I	II	III	IV	V	VI
Material	W	4.7	3.6	3.6	2.8	1.5	3.3
	X	6.1	1.7	1.8	2.6	3.6	4.0
	Y	6.7	6.5	4.2	2.8	2.3	4.9
	Z	3.4	4.9	5.0	4.9	4.5	2.8
	All Observers	5.3	4.3	3.7	3.3	3.0	3.7
Self	W	1.2	1.0	0.5	0.8	0.8	1.4
	X	1.9	1.8	0.5	0.5	0.5	1.5
	Y	0.5	0.6	0.4	0.3	0.8	1.9
	Z	1.6	1.1	0.2	0.5	0.8	1.4
	All Observers	1.3	1.1	0.4	0.5	0.7	1.6
Person	W	4.0	2.4	2.3	0.9	3.0	3.2
	X	2.2	1.7	1.3	2.6	4.0	5.0
	Y	4.6	3.2	2.0	2.4	4.7	3.6
	Z	3.4	3.2	1.7	1.2	3.7	6.3
	All Observers	3.6	2.6	1.8	1.8	3.9	4.6

The actual number of units of *Material* recorded in Film VI was more than four times the number recorded in Film III with the other films intermediate between these two, yet the group indices of disagreement in the two extreme films were exactly the same. The number of units of *Self* recorded in Film IV was almost twice that of the number of units recorded in Film I, and the group index of disagreement is significantly higher in I than in IV. The extreme for *Person* was Film II with more than three times as much of the category recorded as in Films III or V, but the group index of disagreement is probably significantly higher for Film II than for Film III although it is significantly lower than for Film V. Thus no consistent relationship exists, within a given category, between the amount of behavior and the size of the index of disagreement. This suggests the variability of the difficult situations from film to film, irrespective of the actual number of units recorded in a given category.

The *proportion* that a given category assumes in comparison with the other two categories can be similarly examined for any tendency for those films with a relatively high proportion of a category to have a low index of disagreement as compared with those films with a relatively low proportion of the same category. The proportionate amount of behavior recorded in each category in the third observation of each film of Series A for all observers combined, is shown in the following summary table.[1]

Per Cent of Behavior Recorded in Each Category

Film	Material	Self	Person	Total
I	9.9	72.9	17.2	100.0
II	8.4	68.2	23.4	100.0
III	4.7	83.6	11.7	100.0
IV	6.1	76.3	17.6	100.0
V	10.5	80.8	8.7	100.0
VI	13.2	69.9	16.9	100.0

Material is lowest proportionately in Films III and IV and highest in Films V and VI but the group indices of disagreement differ quite insignificantly for these extremes. *Self* is lowest proportionately in II and VI and highest in III and V and there is a significant difference between the indices of these films, low proportions being associated with high group indices. *Person* is low proportionately in V and III and high in II, but no association between the indices and proportionate frequency is found.

It is clear that neither the proportionate relationship of the preferred categories, nor the actual number of units recorded, determines the size of indices of disagreement, and that other variables than a small base or a small proportion of a category must be sought as the main sources of situational difficulty. As a basis from which to operate, therefore, we shall neglect the questions of number and proportion and attempt to differentiate the "hard" from the "easy" films, purely in terms of the variability in the indices of interpretive disagreement on the several films. But turning back to the summary table of these indices, it becomes clear, that even this is impossible, for we cannot make any complete abstraction of the situation from the observers. The individual differences among our observers often dominate over the differences between situations. Since, fur-

[1] These proportions are based on totals for all observers, corrected for half units (a refinement not used in our other analyses unless so noted).

thermore, the differences in indices of disagreement between situations are often statistically insignificant, the best we can do is to define the extremes, i. e., the "easiest" film (or films) for all observers as indicated by the lowest indices of disagreement and the "hardest" film (or films), as indicated by the highest indices of disagreement.

The following summary indicates the "hardest" and "easiest" film for each observer and category. Where two films tied for either place, they are both noted. Where the differences between indices of the "hardest" and another film, or between the "easiest" and another film, are probably insignificant, the other film is noted in parentheses. The differences between the "hardest," on the one hand, and the "easiest," on the other, are probably significant.

"Hardest" Films

	Material	Self	Person
W	I	VI (I)	I
X	I	I (II)	VI
Y	I (II)	VI	V (I)
Z	III (II, IV, V)	I, VI	VI
All Observers	I	VI (I)	VI (V)

"Easiest" Films

	Material	Self	Person
W	V	III	IV
X	II (III)	III, IV, V	III (II)
Y	V (IV)	IV (III, I)	III (IV)
Z	VI (I)	III	IV (III)
All Observers	V (IV)	III (IV)	III, IV

If we can generalize at all here, it seems that I is the "hardest" film for the *Material* category, I and VI for *Self*, and V, VI and I for *Person;* and that no one film is consistently "easiest" for *Material*, III is "easiest" for *Self*, and III and IV for *Person.* I and VI seem to be generally the "hardest," and III and IV generally the "easiest."[1]

What measurable characteristics other than the size of the base can be isolated for the purpose of throwing light on these differences? *A priori*, the only factor in our data capable of

[1] There is some correspondence, but by no means exact agreement, in this analysis of "hard" and "easy" films based on indices of interpretive disagreement and the analysis in Chapter VII, based on percentages of discrepantly timed units. The agreement is complete, with regard to "hard" and "easy" on the *Person* category; partial, i. e., agreeing in regard to the "hard" but not the "easy" on *Self;* but there is no agreement on the *Material* category with regard to the extremes of ease or difficulty.

numerical expression which might be expected to throw light on situational variability is that of the relative duration of each behavior act within a given category. The longer the duration of a single act, the clearer does its essential nature become and presumably the easier it becomes to record. The best[1] measure of duration is a negative one, i. e., the frequency of transitions from a given category to other categories related to a base of the total units of behavior of that category.

A transition index was computed as follows: the number of blocks of behavior in each category was computed, a block being defined as a continuous record of behavior within a given category. The number of breaks in continuity was, obviously, twice the number of blocks (the beginning and end of every block representing a break). The "clean" breaks, i. e., those representing the exit or entrance of the character were excluded from the total, the remainder representing those breaks in a behavior act which followed or were followed by an essential change in the behavior from one of our categories to another. These were computed on the basis of all observers' records for the third observation of each film, were multiplied by 100 and divided by the total units of each category recorded by all observers. The maximum possible number of transitions, so defined, per 100 units would be 200 (if the computation were accurate to the unit) which would occur only if every block of behavior in a given category were a single unit in duration and were always immediately followed and preceded by another category. With the greater refinement used in this computation—accuracy to the half unit—the upper limit could, theoretically, be 400, which would occur only if every block of behavior in a category were a half unit in duration. The following table represents these transition indices for each film.

[1] The reason this is considered "best" is that it is a measure applicable equally well to the real life and to the moving picture observations. In life observations, the behavior of the person observed is classified continuously in one of several categories. In the moving pictures, the character's behavior is non-continuous; he fades from the screen, the scene is shifted to exclude him, etc. Mere duration of an act, therefore, is not directly comparable with our records in life situations, and the number of transitions is a fairer measure than average duration, or average length (in units) of the blocks of behavior.

Transitions per 100 Units
Third Observation, Series A, All Observers, by Categories and Films

Film	Category			
	Material	Self	Person	All categories
I	90	27	75	42
II	53	18	36	26
III	69	7	39	14
IV	88	13	31	21
V	111	21	65	34
VI	74	24	50	35

The "easiest" films (III and IV) have the fewest transitions when all categories are considered together and the "hardest" films (I and VI), the most. The relationship does not, however, hold entirely for the separate categories since V (a relatively easy film) has more transitions than the "hardest" film (I) for *Material; Self* shows a wholly consistent tendency for the "hard" films to be associated with relatively many transitions; and the relationship holds, though not so clearly, for *Person* also.

In order to check this result and to have a somewhat larger body of data with which to work, the films were divided into quarters, and the number of disagreeing units in each category were assigned to the appropriate quarters. For those films where there was one quarter which had twice as many disagreements as another, the former was called the "hard" quarter and the latter the "easy" and the transitions were calculated for each.[1] The following table shows the result of this calculation.

*Transitions per 100 Units for That Quarter (Relative) with the Smallest Number of Disagreements ("easy quarter") and for That Quarter (Relative) with the Largest Number ("hard quarter"), for All Observers Combined, by Category and Film**

Film	Category						All Categories Combined	
	Material		Self		Person			
	"Easy" Quarter	"Hard" Quarter	"Easy" Quarter	"Hard" Quarter	"Easy" Quarter	"Hard" Quarter	"Easy" Quarter	"Hard" Quarter
I	70	134	26	32	74	76	41	48
II	50	84	12	25	26	37	18	31
III	50	89	—	—	16	74	10	22
IV	—	—	—	—	—	—	19	24
V	—	—	—	—	68	87	41	43
VI	86	76	17	33	54	45	27	47

— No "easy" or "hard" quarter for this category and film.

* The "hard" quarter was defined as containing at least twice as many disagreements as the "easy" in proportion to the total units recorded.

[1] If two quarters tied for "easy" or "hard," in a given film, the transitions for the two were averaged.

For all categories combined, the "hard" quarter had regularly more transitions than the "easy" one, but the difference was insignificant in Film V and not impressive in Film I. Of the four films where such computations were possible for *Material*, three showed marked differences between the "hard" with many transitions and the "easy" with relatively few. Film VI, however, showed the opposite tendency, but not to a great degree. *Self* was consistent on all three films where the computation was possible, and *Person* showed a clear tendency in three out of five, an insignificant tendency in one, and the opposite (also probably insignificant) tendency in Film VI.

We may conclude that the relative duration of a behavior act without a transition to another category is probably a factor associated with the hardness or easiness of a given situation. It is obviously, however, not the predominating factor. That our data are not suitable for a more detailed statistical analysis on this point must be clear. It is improbable, indeed, that the factors influencing situational variability are at present capable of numerical expression. Their great importance for the proper sampling of behavior situations with which to test the reliability of our observers, however, made it desirable to attempt to isolate this factor even though it is the only one we can so isolate at present.

This analysis has rather definite implications for determining the "representativeness" of Series C, which, it will be remembered, was used as a basis for estimating the observers' tendencies to distort reality. Series C was patently unrepresentative in one respect. There were more units of *Material* recorded than of *Self*, and only about twice as many of *Self* as of *Person*. Thus there was a preponderance of the preferred categories, whereas all our other films showed an overwhelming preponderance of self behavior. The table on the next page indicates the transition index separately for each of the characters used in our comparison in the Series C normal speed observations.

If transitions be taken as an index of the difficulty of an observational situation, Series C is clearly much harder than any of the films of Series A. The *Material* category is on about the same level of difficulty but *Self* and *Person* are considerably more difficult. This may be taken as an implication that our slow

Transitions per 100 Units, Series C, normal speed. All Observers, by Character Observed and Category

Character Observed	Category			All Categories Combined
	Material	Self	Person	
Philip	79	169	174	119
John	78	157	163	113
Billy	94	133	223	125

speed observations in Series C were very far from a refined approximation to reality, because of the inherent difficulty of the situations represented in this film. The general tendencies we have found are probably true; the proportions representing these tendencies would probably have varied considerably with other observational situations.

CHAPTER XI

CONCLUSIONS

We have dealt, in this volume, with the observability of certain aspects of social interaction. In Part I, we described a technique for recording overt behavior indicative of the individual's reactions to the members of the group in which he is involved, and analyzed the results of the application of this technique to several very different sorts of groups. Our general hypothesis has been that simple overt acts, capable of precise definition, can be selected in such a way as to be significant indicators of more complex relationships. Specifically, we have assumed that the interrelationship between the mere number, range and duration of contacts individuals make with other persons, the duration and extent of their manipulations of the material environment and the duration and extent of their periods of withdrawal from active contact with persons or from manipulation of materials are significant personality differentia. We have further assumed that group patterns and individual deviates can be determined in such a way that differences between groups in various situations, as well as changes in a given group and in the individuals comprising it, may be studied, provided only that the precision of the technique we have devised be of a known and adequate, degree.

Our observational technique, as applied to the earliest age levels, is entirely in terms of behavior acts: physical and verbal contacts with persons, defined as social behavior acts; manipulation of materials, defined as material acts; laughing, crying and other physical activity unconnected with materials, defined as egoistic acts; and absence of overt activity, defined as passive egoistic acts. At the kindergarten level, our observational technique includes the concept of function, and we attempted to classify each act, defined in the same terms as at the earlier level, in terms of its connection with a predetermined job or its lack of such connection. At the later levels, this separation of acts on a functional basis (i. e., acts determined by the structure of the situation and acts not so determined) is carried even fur-

ther—a modification which we consider to be of fundamental importance, if observational results in structuralized situations are to have any significance as personality differentia.

The application of this technique of classifying behavior acts, as they occur, in terms of their social, material and self components was made possible by the use of a continuous time scale divided into small time units, i. e., by neglecting the determination of *numbers* of discrete acts and emphasizing the *duration* of each of the components of these acts. This also made possible a systematic weighting of the components and a simple treatment of the results as additive time units.

The result of the application of this technique was highly satisfactory, especially at the nursery school and kindergarten age levels, in defining group behavior patterns and individual deviates and in the determination of stability or change in the patterns over a period of time. At the nursery school level, furthermore, the simple indices derived by this technique were "validated" by a series of independent studies (by other observational techniques) of the social interaction of the same groups of children.

Since this technique was designed for rather wide application, especially in studying the range of social interaction in different sorts of child-groups, and since its application produced results of apparently great precision, it seemed essential to study in some detail the real limits of this precision. If such a study were to be adequate, it could not stop at the naïve stage of reasoning where it is assumed that adequate precision has been obtained if the records of any given situation made by two observers are highly correlated, or the equally naïve stage where it is assumed that precision has been achieved if group or individual stability can be "predicted" within certain limits upon increasing the amount of data by certain proportions. It was essential to examine the observer as a scientific instrument, to the variability of which a large number of factors contribute. This instrument had to be examined for its agreement with other similar instruments measuring the same materials, its agreement with itself measuring the same materials consecutively, its systematic errors when compared with a more accurate instrument and its variations in efficiency in situations of different degrees of complexity.

Part II represents an attempt to make such an examination of

the functioning of the observer with a technique such as we have used in our several studies. In order to overcome the practical limitations of making such an examination in ordinary life situations, with ever changing materials, we made series of observations of the behavior of the characters in motion picture films. We used a technique strictly comparable with those used in our other studies, the observer being required to differentiate the observed behavior in terms of three mutually exclusive categories, the difference between one of these categories and the other two being entirely in terms of overt activities, and the difference between another of these categories and the other two depending on differences in the functions of overt activities.

We recognized that a technique depending on the assignment of behavior items to time units introduces by its very structure errors in the observational record. Our attention was, therefore, first directed toward the measurement of the amount and direction of our observers' disagreements which could be attributed to failure to assign behavior to the same point along the time scale. This measurement had to be on a relative basis and attempted to determine the consistency in amount and direction of each observer's timing disagreements with each of the other observers on simultaneous records.

Next, the proportion of each category interpreted by each observer in disagreement with each other observer on simultaneous records was measured in relation to his total agreements on the same category.

Third, an attempt to study self consistency on successive observations of the same event was made by inference only, i. e., we measured progressive tendencies of the observers to change in their proportions of disagreements with each other on consecutive observations.

Fourth, we attempted to describe and measure the various sorts of systematic errors made by our observers, i. e., to determine whether on many sample-observations the timing tendency of each observer was consistently in a particular direction; whether in his interpretive disagreements with other observers he always "favored" a given category at the expense of another; whether, in his own consecutive observations, a "favoring" tendency developing in the beginning of an observation tended

to persist throughout; and, finally, we attempted to determine the extent to which our observers distorted "reality" in their records by consistent underemphasis of certain categories and overemphasis of others.

Fifth, we attempted to define and measure certain variables in the several situations observed which might contribute to the varying efficiency of the technique.

This whole experimental study of the observers and the technique must be taken more as a demonstration of the fact of variability than as a determination of its degree, or an isolation and exact measurement of the contribution of each factor to the total variability. Our statistical basis is by no means secure. Four observers do not constitute a stable observational norm. Three consecutive observations are not a large enough number to permit the isolation of consistency from reliability. Our determination of "reality" was necessarily crude. We have tried not to overemphasize the importance of our numerical results as such. At the same time, our analysis is made in considerable quantitative detail, because the experiment was carried through with great care. The results, therefore, should provide a not unimportant basis upon which to build further.

We have shown quite definitely for our small group of observers, with a rather extensive series of observational samples, that systematic errors of observation are an important factor in variability; that the observational error varies with the situation observed; that observer-inconsistency probably plays a definite rôle; that, with our technique, timing and interpretation errors are at present almost inextricably interrelated; and that distortion of "reality" quite clearly occurs in observational records. If we have done nothing more than demonstrate that precision cannot be determined by a simple measure of agreement between two equally trained observers, we have made progress. That we have raised more problems than we have been able to solve is no reason for discouragement as to the future of observational techniques in sociological studies. It is impossible to make real progress in the use of an instrument until its nature is thoroughly understood.

APPENDICES

APPENDIX A

Transcription of Records

A detailed description of the procedure followed in transferring the records of the four observers from the original forms to master sheets is given below. Chart A shows the simultaneous records made by the four observers, W, X, Y and Z, over a period of 15 five-second intervals in the third observation of the film *This Reckless Age* (the first fourth of page 3 of the record). Chart B shows the corresponding column of the master sheet containing these four transcribed records.

The rules for transcription of the records were as follows:

(1) Number the master sheet pages to correspond with those of the original record.

(2) Put the title of the picture, name of character observed and number of observation on each sheet. Show on the first sheet the date and time of the observation and the number of the theatre seat occupied by each observer if this is shown on the original records.

(3) Use one column of the master sheet for each observer.

(4) Use a green pencil for *Material* (dotted line in Chart B), a red pencil for *Self* (straight line in Chart B), and a blue pencil for *Person* (x-x- line in Chart B).

(5) Copy records of *Material* at the left of the column, those of *Self* in the center, and those of *Person* at the right.

(6) Any record appearing at all in any half interval is to be copied as a full half interval of behavior.

(7) When the recorded behavior does not meet a following exit line, consider the exit line the more accurate and continue the record of behavior to meet it.

(8) When there is an absence of recorded behavior lasting less than one half interval and there is no exit line, continue the preceding record of behavior to meet the beginning of the following.

(9) Where there is an absence of recorded behavior lasting one half interval or more and there is no exit line, consider that the line was unintentionally omitted and supply it where the preceding behavior ends.

Chart A. Sample Sections of the Records Made by Observers W, X, Y and Z of the Motion Picture Character *Frances Dee* in the Film *This Reckless Age.*

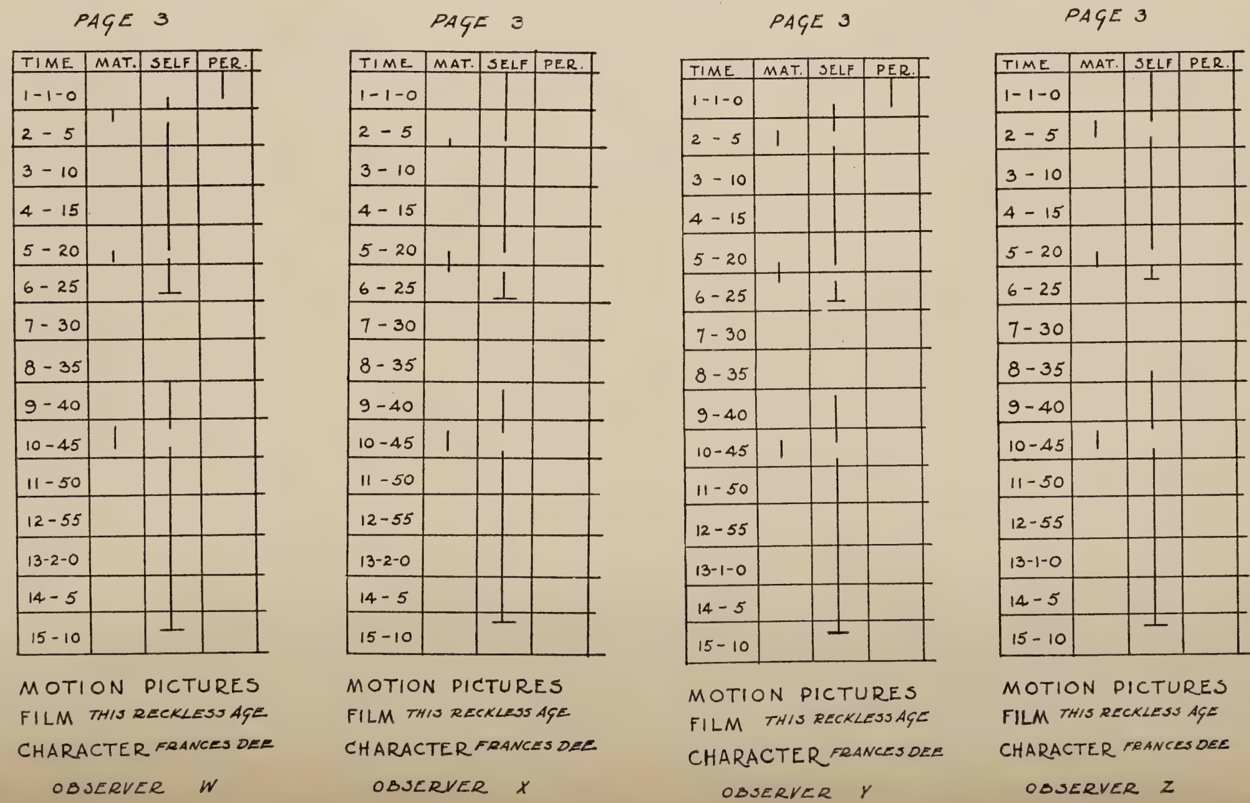
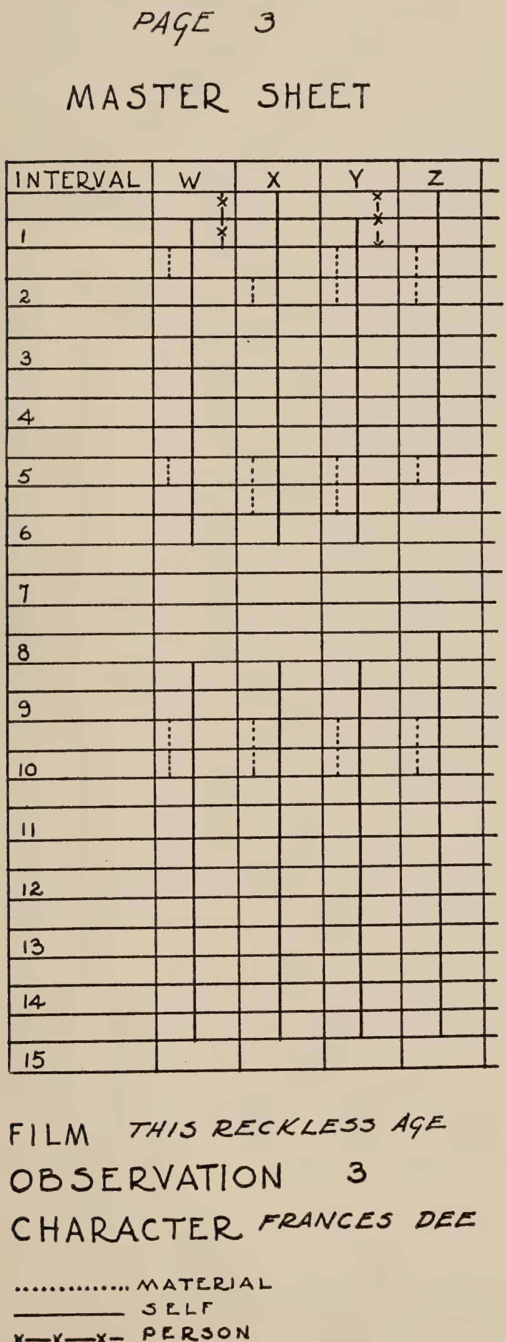


Chart B. Sample Section of a Master Sheet Illustrating the Method of Transcription as Applied to the Original Records Shown in Chart A.



(10) Where the end of a record in one category of behavior overlaps the beginning of the record in another, consider the beginning of the second as the more accurate and stop the first at the point where the second begins.

(11) Make no changes on the original records but make a note of situations where judgment is required, as in the case of supplying an exit line, and consult the recorder for final decision.

Rules 6 to 10 inclusive were designed to make the reproduction of the records as accurate as possible. They depended upon a condition inherent in the method, namely, that the most accurate record possible of a change in behavior, whether it was an exit or a change of category, was made by glancing at the clock at the moment the change was observed and recording the change as occurring at the time noted. Thus, the beginning of the new record should have been fairly accurate even if the end of the preceding one was not.

With the exception that the quarter interval was used as the unit of transcription instead of the half interval, the same rules were used for Series C as for Series A. The quarter interval was used instead of the half interval to determine when an exit line should be supplied if none was shown on the original record. Thus an absence of behavior lasting less than a quarter interval was treated as no exit, whereas an absence of behavior lasting a quarter interval or more was interpreted as being due to an exit and the line was supplied. One additional rule was applied to records of Series C. Whenever there was a break in the behavior within a category due to a brief exit or to the brief occurrence of behavior in another category, and this break was not indicated by a vacant quarter interval in that category on the master sheet, the break was shown by a short horizontal line.

APPENDIX B

Effect of Inadequately Synchronized Timing Instruments on Observer Reliability

In a series of observations preceding those described in this monograph (i. e., made during 1930–1931) a serious cause of disagreement among observers was found to be associated with discrepancy in timing due to inadequately synchronized stop-watches. It should be remembered that, in this preliminary motion picture study, the records were of the type shown in Form B of the Arrington technique,[1] i. e., they were non-continuous records of discrete occurrences of language and physical contacts. They are not directly comparable, therefore, with the *Material-Self-Person* records discussed in Part II of this monograph. The factor of discrepancy in timing instruments had not been important in previous studies of behavior in life situations, since stop-watches had been run for very brief periods at a time,—five-minute, ten-minute, or at most, fifteen-minute periods. When, in this study, stop-watches were run continuously for periods of an hour to an hour and a half, it soon became evident that a lack of synchronization between watches used by different observers was causing marked deviations in the timing of the same events. Watch readings were, therefore, taken at the end of each observation in order that the effect of watch differences might subsequently be determined.

To determine the extent of deviation in records associated with discrepancy in timing instruments, the tabulated data per observation and pair of observers were arranged by amount and direction (retardation or acceleration) of watch differences. If, for example, Observer W's watch reading at the end of one and a half hours of recording was 54.6″, and Observer X's 54.2″, a difference of .4″ was recorded as acceleration for W and as retardation for X. The effect of an accelerated watch was to cause retarded records, that of a retarded watch, to cause accelerated records, since the fast watch makes the record appear

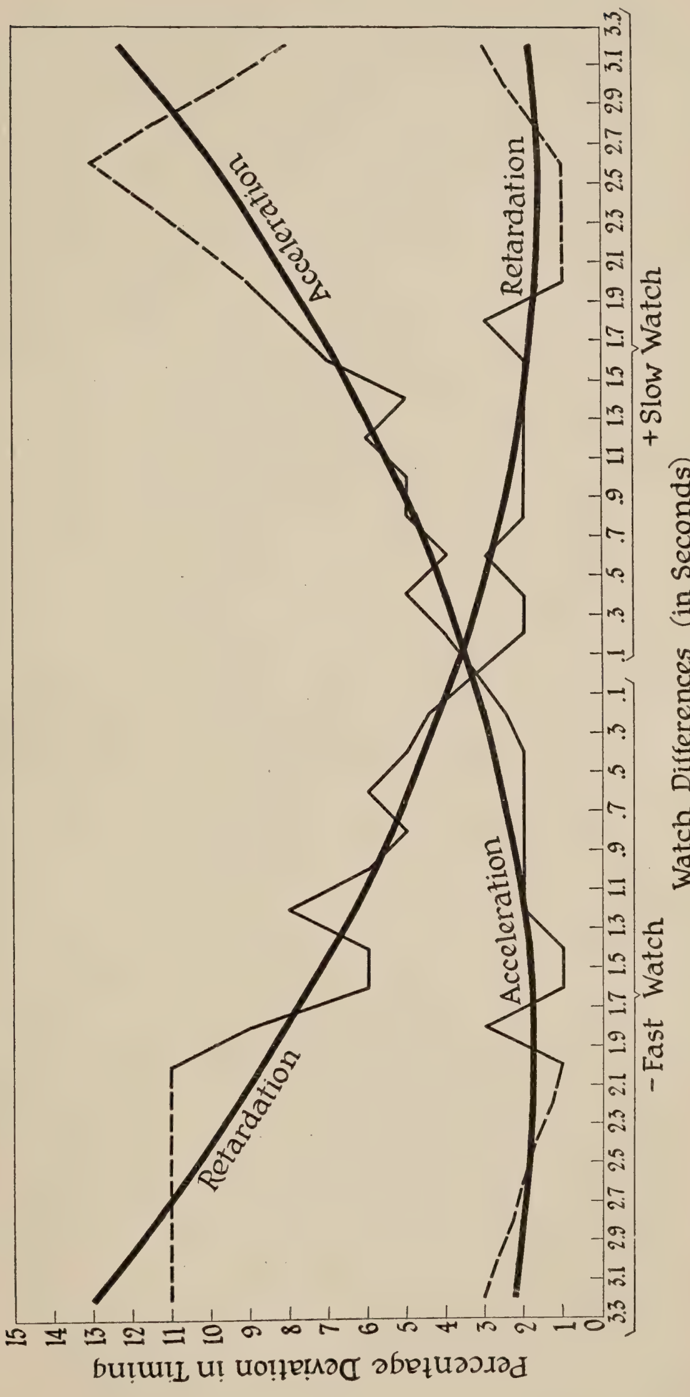
[1] See p. 3, *supra*.

in a later five-second interval, the slow watch, in an earlier interval. Percentages of accelerated and retarded intervals were computed for watch differences ranging from 0 to 3.3 seconds, the latter being the largest amount of difference noted. The records of the observer having the fast watch and of the one having the slow watch in each pair were tabulated separately, each observer's record being compared with the other observer's record as "standard." An accelerated interval was defined as an uncorroborated interval preceding and adjacent to an interval in which the same behavior was recorded by the "standard" observer; a retarded interval, as an interval following and adjacent to an interval in which the same behavior was recorded by the "standard" observer. Uncorroborated intervals, both preceded and followed by intervals of the same behavior by the "standard" observer were defined as indeterminate timing and were counted as half acceleration and half retardation. The data thus tabulated were plotted according to amount and direction of watch difference and second-degree parabolas were fitted to the data. Retardation and acceleration trends in terms of amount and direction of watch difference for the three pairs of observers combined are shown in Chart C.

The trends of acceleration and retardation, as shown here, indicate that about 2% of the intervals were accelerated and retarded "normally" regardless of watch differences and individual timing biases, i. e., observers tended to record about 2% of their data in the interval immediately preceding that of the "standard" observer and about 2% in the interval immediately following. The observer whose tendency was to retard his records would obviously have a higher percentage of retardation and the one who tended to have accelerated records in comparison with other observers would have a higher percentage of acceleration, without regard to watch differences. With a watch difference of one second, we would expect, on the basis of these trends, an additional percentage of approximately 3% of the intervals to be retarded by the observer whose watch was fast and approximately 3% to be accelerated by the observer whose watch was slow. When the watch difference was two seconds, the percentage of expected deviation due to watches would increase to 7%; and for a three second difference in

CHART C

Retardation and Acceleration Trends of Three Pairs of Observers in Terms of Amount and Direction of Watch Difference

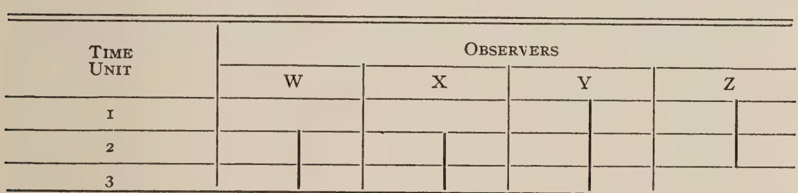


watches the expected retardation and acceleration attributable to timing instruments would be 10%. It is obvious, from these findings, that very exact synchronization of time pieces used simultaneously by different persons over appreciable periods of time is essential if records are to be taken continuously. It is also essential that watches should be started and stopped at exactly the same instant. The fineness of the time unit used in this study obviously made even minor deviations in the timing instruments a factor tending to obscure timing differences due to the personal idiosyncrasies of the observer. A difference of a single second might cause an event to be recorded in one interval by one observer and in an earlier or later interval by another. Fortunately, in view of its implications for recording in the life situation, there appeared to be no significant effect of the timing deviations due to stop-watches upon the total record of behavior.

APPENDIX C

Calculation of Percentage Indices of Timing Tendency

The method of tabulating and weighting all recorded units of behavior for a given category and of calculating timing tendency indices is illustrated below. A simple example may serve to clarify the initial process of tabulation. Unit 1, in the accompanying hypothetical record, would be tabulated on Form A[1] in the horizontal row labeled YZ, once in the "No Record" + space, since there is no record in the preceding time unit, and once in the WXYZ − space, since all four observers record in the following time unit. Unit 2 would be tabulated in the horizontal row labeled WXYZ, once in the YZ + space and once in the WXY − space.



Unit 3 would be recorded in the horizontal row labeled WXY, once in the WXYZ + space and once in the "No Record" − space.

Form A shows the tabulation of all units of *Material* recorded by the four observers in the third observation of the film *This Reckless Age* in terms of the particular combinations of observers recording the behavior synchronously, i. e., in the same time unit, and the particular combinations recording in the preceding and in the following units. The horizontal rows represent the so-called "synchronous" units, the vertical + columns, the retarded, and the vertical − columns, the accelerated, units, each unit being tabulated twice. Thus, reading the top line of this form, we see that, out of a total 41 (82/2) units of *Material* recorded synchronously by the four observers, 21 were preceded

[1] See Form A, *infra*.

FORM A

THIS RECKLESS AGE Observation 3 Material

	WXYZ		WXY		XYZ		WYZ		WXZ		WX		WY		XY		XZ		YZ		WZ		W		X		Y		Z		No Record		Total
	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	
WXYZ	𝍸𝍸𝍸𝍸/	𝍸𝍸𝍸𝍸/		////	/		//	///					/	//		///	//		/	//	///				//	/			///		𝍸	𝍸	82
WXY	////		/	/			///	/				/		/	/	//										/						//	18
XYZ		/			/	/	//	//	/																	/					//	/	12
WYZ	///	//	/	///	//	//	/	/								//		/	/				/				/	//	/	/	𝍸	//	32
WXZ						/														/									/		/		4
WX			/																													/	2
WY	//	/	/										/	/		/				/								/	/		//	//	14
XY	///		//	/			//						/		/	/										///	/					𝍸	20
XZ		//					/																	/							//		6
YZ	//	/						/	/				/						/	/								/			//	///	14
WZ		///																							/						//		6
W								/									/						//	//							///	///	12
X	/	//	/		/										///							/			//	//		/	/		𝍸	𝍸///	28
Y							//	/					/			/			/						/		//	//			/	////	16
Z		///					/	/		/				/												/			/	/	𝍸//	/	18
Total	36	36	7	9	5	4	14	11	2	1	0	1	5	5	5	10	3	1	4	5	3	1	3	3	6	9	4	7	8	2	37	37	284

by the combination WXYZ, one by XYZ, two by WYZ, one by WY, two by XZ, one by YZ, three by WZ, two by X, three by Z, and five by no record. Of these same units, 21 were followed by the combination WXYZ, four by WXY, three by WYZ, two by WY, three by XY, two by YZ, one by X, and five by no record.

From Form A were derived tabulations for each of the six pairs of observers comparable to the one shown below for the pair WX (Form B). All entries on Form A in which W or X

FORM B

THIS RECKLESS AGE *Observation 3* *Material* *Observers* *WX*

Observers	WX		W		X		No Record		Total
	+	−	+	−	+	−	+	−	
WX	21 0 4 1 0 1	21 4 0 1 1 0	2 1 3 3 0	3 2 0 1 1	1 0 2 2 1 0 0	0 3 0 1 2 1 1	1 3 5 0 0 1 1 0	2 0 5 2 1 0 0 1	
Total	27	27	9	7	6	8	11	11	106
W	3 1 2 1 0	2 3 1 0 3	1 1 1 0 2	1 0 1 1 2	2 0 0 0 1 1	2 2 1 1 0 0	1 1 1 5 0 0 1 2 2 3	0 2 1 2 1 1 0 2 0 3	
Total	7	9	5	5	4	6	16	12	64
X	0 1 3 2 0 1 1	1 0 0 1 2 2 0	2 2 1 1 0 0	2 0 0 0 1 1	1 0 1 0 1 3 2	1 1 1 3 0 0 2	2 1 0 2 0 1 5	1 0 5 0 1 0 8	
Total	8	6	6	4	8	8	11	15	66
Total	42	42	20	16	18	22	38	38	236

appeared were transferred to Form B, the records of other observers being disregarded. All entries in the horizontal rows labeled WXYZ, WXY, WXZ and WX were recorded in the horizontal WX space of Form B under WX, if derived from the vertical columns bearing the same labels; under W, if derived from vertical columns labeled WYZ, WY, WZ or W; and under

X, if derived from the XYZ, XY, XZ or X columns. In the horizontal W space of Form B were recorded all entries in the horizontal rows involving W without X, and in the X space, all entries in horizontal rows involving X without W. Interpreting Form B for the pair WX, we find that these observers had a total of 53 "synchronous" units (106/2), that W had 32 (64/2) units uncorroborated by X, and that X had 33 (66/2) units uncorroborated by W. Of the 53 "synchronous" units, 27 were preceded by units in which both observers recorded; 9, by units in which W alone recorded; 6, by units in which X alone recorded; and 11, by no record by either observer. Similarly, 27 of the 53 units were followed by units in which both observers recorded *Material;* 7, by units in which W alone recorded; 8, by units in which X alone recorded; and 11, by no record by either observer. The horizontal total rows for WX, W and X in Form B were weighted by the number of persons recording in terms of "synchronous," retarded and accelerated timing, and the difference between the total weighted retarded (+) and total weighted accelerated (−) units for each observer was divided by the average number of "synchronous" units per observer to get the percentage index of timing tendency.

The calculation follows:

WX				W				X			
Syn.	Ret.	Syn.	Acc.	Syn.	Ret.	Syn.	Acc.	Syn.	Ret.	Syn.	Acc.
54	54	54	54	7	14	9	18	8	16	6	12
18	9	14	7	5	5	5	5	6	6	4	4
12	6	16	8	4	4	6	6	8	8	8	8
22	0	22	0	16	0	12	0	11	0	15	0
106	69	106	69	32	23	32	29	33	30	33	24

W			X			
Syn.	Ret.	Acc.	Syn.	Ret.	Acc.	
106	69	69	106	69	69	$\begin{matrix} W - \\ X + \end{matrix} \frac{6}{138.5} = \pm 4.3\%$
32	23	29	33	30	24	
138	92	98	139	99	93	

All pairwise indices in Table XI (Chapter VII) were computed in this manner. Indices in Table XII (Chapter VII) were

obtained by totaling the weighted "synchronous," retarded (+), and accelerated (−) units per observer in three pairwise combinations with other observers and dividing the difference between the weighted + and weighted − units (data for three pairings combined) by the average number of weighted "synchronous" units per observer (data for three pairings combined). This process is shown below for the third observation of *This Reckless Age, Material.*

Pair of Observers	W			X			Y			Z		
	Syn.	Ret.	Acc.	Syn.	Ret.	Acc.	Syn.	Ret.	Acc.	Syn.	Ret.	Acc.
WX	138	92	98	139	99	93	—	—	—	—	—	—
WY	158	97	111	—	—	—	177	134	120	—	—	—
WZ	147	101	90	—	—	—	—	—	—	149	95	106
XY	—	—	—	152	105	102	170	124	127	—	—	—
XZ	—	—	—	138	103	87	—	—	—	139	91	107
YZ	—	—	—	—	—	—	174	138	115	157	97	120
	443	290	299	429	307	282	521	396	362	445	283	333

$$W - \frac{9}{460} = - \ 1.9\%$$

443
429
521
445
$4 \overline{)1838}$
$459.5 = 460$

$$X + \frac{25}{460} = + \ 5.4\%$$

$$Y + \frac{34}{460} = + \ 7.4\%$$

$$Z - \frac{50}{460} = - \ 10.9\%$$

$$WXYZ \pm \frac{59}{460} = \pm \ 12.8\%$$

APPENDIX D

Calculation of Percentage Indices of Interpretive Disagreement

The calculation of a crude index of interpretive disagreement for each observer in each pairwise combination with other observers and for each observer compared with three other observers involved the following steps.

(1) Tabulation, from the master sheets described in Appendix A, of units of disagreement[1] on each pair of categories without regard to the record of the third category, as, for example, *Material-Self*, W—XYZ (W having a record of *Material* when the other three observers record *Self*),

(2) Summarization, for every possible combination of observers, of disagreements in recording each pair of categories,

(3) Selection, from this general summary, of disagreements for each pair of observers,

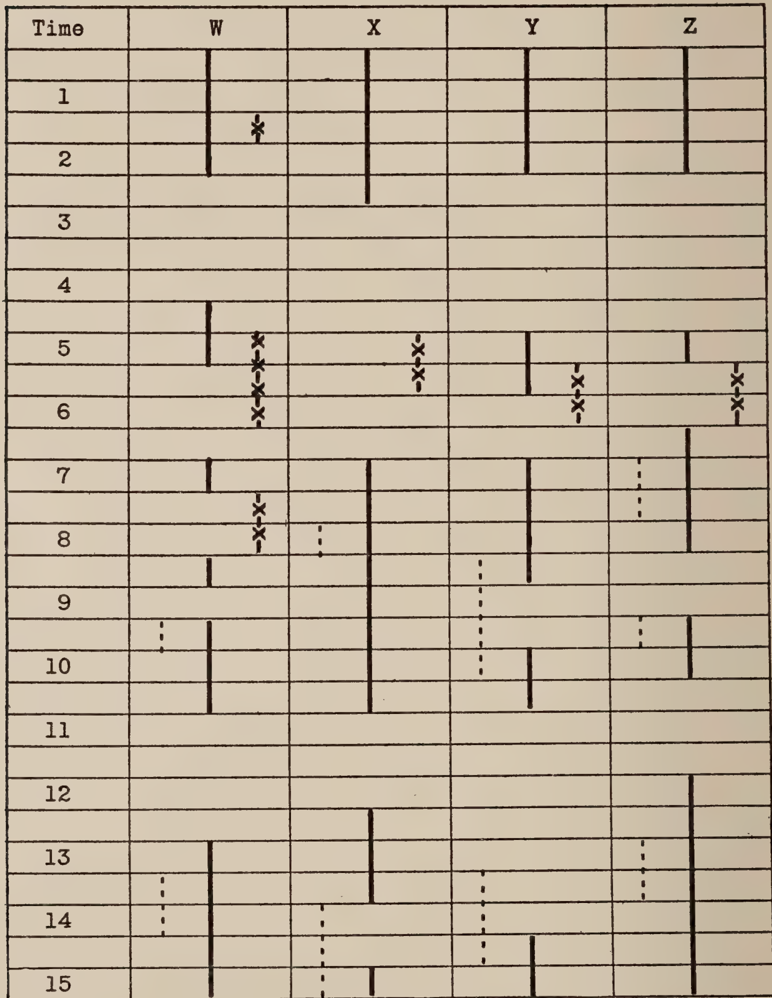
(4) Assembling, by observer and category, of total weighted units (derived from the timing analysis) and disagreements on each pair of categories, and calculation of the pairwise index representing the ratio between total units and disagreements for each observer by category, film and observation,

(5) Combination of all relevant data necessary for the calculation of average indices per film and observation shown in Table XVI and all indices in Table XVII (Chapter VIII).

It was necessary to tabulate disagreements for only three of the six pairs of categories, since the data for the other three combinations could be obtained indirectly. For example, if the *Material-Self* tabulation shows that Observer W recorded *Material* in ten units when Observers X, Y and Z recorded *Self*, we know, by reversing the categories and observers, that Observers X, Y and Z recorded *Self* in ten units when Observer W recorded *Material*. Disagreements were tabulated for the *Material-Self*, *Material-*

[1] Only disagreements lasting a whole unit, i. e., 2½″, for both observers compared were considered in this analysis. Thus, if one observer had a record of both *Material* and *Person* and the other observer had a record only of *Material* within a given time unit, the unit was not tabulated as a disagreement.

CHART D. MASTER SHEET SHOWING INTERPRETIVE DISAGREEMENTS OF FOUR OBSERVERS



FILM ____ This Reckless Age ____

OBSERVATION ____ 3 ____

- - - - - - - - - Material

———— Self

x-x-x-x-x-x Person

Person and *Self-Person* combinations and were obtained by transposition for the *Self-Material, Person-Material* and *Person-Self* combinations.

As an illustration of the tabulation of disagreements, the first fourth of Page 4 of the master sheet for the third observation of the film *This Reckless Age* is shown in Chart D. Reading the four columns of this section of the master sheet for disagreements in the recording of *Material* and *Self*, and disregarding partial disagreements as defined above, we find the first discrepancy of a whole unit in the second half of the ninth interval, "lower 9," where Y recorded *Material* and X *Self*. This is tabulated as Y-X under *Material-Self*. The same disagreement appears in the succeeding unit, "upper 10," and also in "upper 14" but, in "lower 14," X and Y record *Material* and Z *Self*. Continuing this process for the three pairs of categories, we find that the total interpretive disagreements in this section of the record are as follows:

Material-Self		*Material-Person*		*Self-Person*	
Lower 9	Y—X	Upper 8	Z—W	Lower 5	YZ—X
Upper 10	Y—X	Upper 9	X—W	Upper 8	XYZ—W
Upper 14	Y—X			Lower 8	XYZ—W
Lower 14	XY—Z				
Upper 15	X—WZ				

From these tabulations of disagreements on each pair of categories were assembled the units occurring for each combination of disagreeing observers. Form C below shows all the interpretive disagreements occurring in Film I, Observation 3. From this summary by film and observation the disagreements of each observer with each other observer were isolated for each pair of categories. For example, the units in which W recorded one category in opposition to the synchronous records of X are found in the combinations 3 (WYZ—X), 6 (WY—XZ), 10, 13, 21, 23, 35, 36 and 47. For the film-observation in our example, (Chart D) these disagreements of W with X total 8 for *Material-Self* and 2 for *Material-Person*. The total units of any category recorded by X when W recorded a different category are found in the combinations 2, 7, 8, 15, 17, 26, 38, 40 and 48. For *Material-Self* and *Material-Person* in the example, the totals

FORM C

Interpretation Analysis—Summary of Disagreements
Film, This Reckless Age. Observation 3

Observers	Category M-S	Category M-P	Category S-P	Observers	Category M-S	Category M-P	Category S-P
1. WXY-Z			6	23. W-X	2		
2. XYZ-W	1		3	24. W-Y			
3. WYZ-X				25. W-Z	1		
4. WXZ-Y				26. X-W		1	2
				27. X-Y			4
5. WX-YZ			3	28. X-Z			
6. WY-XZ	2			29. Y-W			
7. XY-WZ	2			30. Y-X	4		
8. XZ-WY			2	31. Y-Z	1	1	1
9. YZ-WX	1			32. Z-W		1	1
10. WZ-XY				33. Z-X		1	
				34. Z-Y	1		1
11. WX-Y			2				
12. WX-Z				35. W-XY		1	
13. WY-X	3		1	36. W-XZ			
14. WY-Z	3			37. W-YZ			
15. XY-W			1	38. X-WY	1		3
16. XY-Z	1	2		39. X-YZ	1		2
17. XZ-W			1	40. X-WZ	2		
18. XZ-Y				41. Y-WX			
19. YZ-W				42. Y-XZ	1		
20. YZ-X	2		2	43. Y-WZ	1		
21. WZ-X			1	44. Z-WX			
22. WZ-Y			1	45. Z-WY	1	1	3
				46. Z-XY			1
				47. W-XYZ	1	1	1
				48. X-WYZ		3	
				49. Y-WXZ	1		
				50. Z-WXY	1		

are 6 and 4, respectively. Then, of course, we know that X recorded *Self* in 8 units and *Person* in 2 units when W recorded *Material* and also that W recorded *Self* and *Person* in 6 and 4 units, respectively, in disagreement with a record of *Material* by X. Thus, either by selection from Form C or by transposition of data so selected, we found the number of disagreements of each observer with each other observer in each pair of categories.

The numbers of total weighted units recorded by each observer in each category, that is, the bases used in the calculation of percentages of disagreement were the numbers of "synchronous" units in the timing analysis. This base in a pairwise analysis represents the total units weighted by either 2 or 1 according to the number of observers recording synchronously. It may, therefore, vary for one observer in different pairings with other observers. The data relating to total units and disagreements for W and X in the interpretation of *Material*, in Film 1, Observation 3, and the resulting indices of disagreement are:

	Total Weighted Units	Disagreements		
		Material-Self	Material-Person	Material-Self and Person
W paired with X	138	8	2	10
X paired with W	139	6	4	10

The percentages of disagreement shown in the pairwise indices for these observers are for W, 7.2% (10/138), and for X, 7.2% (10/139). Group indices for individual observers were computed by summation of the observer's total units and his disagreements with each other observer. The group indices of W and of X in the interpretation of *Material*, Film 1, Observation 3, were obtained as follows:

	Total Weighted Units	Total Disagreements
W—X	138	10
—Y	158	3
—Z	147	8
— Group	443	21
X—W	139	10
—Y	152	5
—Z	138	11
— Group	429	26

The percentage group indices of disagreement of these observers are 4.7 (21/443) for W, and 6.1 (26/429) for X.

The index of disagreement for the group was found in a comparable manner by totaling the relevant units of each observer. This group measure of unreliability in the interpretation of *Material* in Film I, Observation 3, is found, by the following simple tabulation, to be 5.3%.

	Total Weighted Units	Total Disagreements
W	443	21
X	429	26
Y	521	35
Z	445	15
Group	1838	97

$$\frac{97}{1838} = 5.3\%$$

APPENDIX E

Standard Deviations of Percentages

Two reference tables giving the standard deviations of percentages from 0.5 to 20.0 computed on bases of varying sizes are presented here. Table I shows for each category the standard deviations of percentages computed on the average pairwise base per observer, observation and film, and the average base per observer for the six films combined. Table II shows the standard deviations of percentages computed on the corresponding bases per observer, when the data for three separate pairings of each observer with each other observer are totaled. These average bases are, obviously, three times as large as those in Table I. These tables were used in Chapters VII and VIII as a basis for estimating roughly the possible significance of percentages obtained. Whenever the size of base deviated too greatly from the bases shown in the tables to warrant an estimate by interpolation, the standard deviations were computed on the appropriate bases. The formula used in all cases was

$$\sigma_p = \sqrt{\frac{PQ}{N}}$$

where P = observed percentage

Q = 100 minus observed percentage

N = number of cases upon which the per cent is based.

TABLE I

Average Number of Units in Pairwise Bases Used in Computing Percentages of Timing Tendency Per Observer and Observation in Series A and Standard Deviations of Percentages of These Bases, by Film and Category

Category	Film	Average Number of Units in Pairwise Base by Film	Standard Deviations of Each of the Following Percentages on the Corresponding Base																									
			%	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	6.0	7.0	8.0	9.0	10.0	11.0	12.0	13.0	14.0	15.0	16.0	17.0	18.0	19.0	20.0
Material	I	439	σ	0.3	0.5	0.6	0.7	0.7	0.8	0.9	0.9	1.0	1.0	1.1	1.2	1.3	1.4	1.4	1.5	1.6	1.6	1.7	1.7	1.8	1.8	1.8	1.9	1.9
	II	590	σ	0.3	0.4	0.5	0.6	0.6	0.7	0.8	0.8	0.9	0.9	1.0	1.1	1.1	1.2	1.2	1.3	1.3	1.4	1.4	1.5	1.5	1.5	1.6	1.6	1.6
	III	222	σ	0.5	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0	2.1	2.2	2.3	2.3	2.4	2.5	2.5	2.6	2.6	2.7
	IV	530	σ	0.3	0.4	0.5	0.6	0.7	0.7	0.8	0.9	0.9	0.9	1.0	1.1	1.2	1.2	1.3	1.4	1.4	1.5	1.5	1.6	1.6	1.6	1.7	1.7	1.7
	V	659	σ	0.3	0.4	0.5	0.5	0.6	0.7	0.7	0.8	0.8	0.8	0.9	1.0	1.0	1.1	1.2	1.2	1.3	1.3	1.3	1.4	1.4	1.5	1.5	1.5	1.5
	VI	924	σ	0.2	0.3	0.4	0.5	0.5	0.6	0.6	0.6	0.7	0.7	0.8	0.8	0.9	0.9	1.0	1.0	1.1	1.1	1.1	1.2	1.2	1.2	1.3	1.3	1.3
	I–VI	3364	σ	0.1	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.4	0.5	0.5	0.5	0.5	0.6	0.6	0.6	0.6	0.6	0.6	0.7	0.7	0.7
Self	I	2964	σ	0.1	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.5	0.5	0.5	0.5	0.5	0.6	0.6	0.6	0.6	0.6	0.6	0.7	0.7	0.7
	II	4957	σ	0.1	0.1	0.2	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.4	0.4	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.6	0.6
	III	3568	σ	0.1	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.5	0.5	0.5	0.5	0.5	0.6	0.6	0.6	0.6	0.6	0.6	0.7	0.7	0.7
	IV	5847	σ	0.1	0.1	0.2	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.4	0.4	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.6	0.6
	V	4259	σ	0.1	0.1	0.2	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.4	0.4	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.6	0.6
	VI	4939	σ	0.1	0.1	0.2	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.4	0.4	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.6	0.6
	I–VI	26534	σ	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Person	I	789	σ	0.2	0.4	0.4	0.5	0.6	0.6	0.7	0.7	0.7	0.8	0.9	0.9	1.0	1.0	1.1	1.1	1.2	1.2	1.3	1.3	1.3	1.4	1.4	1.4	1.4
	II	1733	σ	0.2	0.2	0.3	0.3	0.4	0.4	0.4	0.5	0.5	0.5	0.6	0.6	0.7	0.7	0.7	0.8	0.8	0.8	0.8	0.9	0.9	0.9	0.9	1.0	1.0
	III	555	σ	0.3	0.4	0.5	0.6	0.7	0.7	0.8	0.8	0.9	0.9	1.0	1.1	1.1	1.2	1.3	1.3	1.4	1.4	1.5	1.5	1.6	1.6	1.6	1.7	1.7
	IV	1426	σ	0.2	0.3	0.3	0.4	0.4	0.4	0.5	0.5	0.5	0.6	0.6	0.7	0.7	0.8	0.8	0.8	0.9	0.9	0.9	0.9	1.0	1.0	1.0	1.0	1.1
	V	514	σ	0.3	0.4	0.5	0.6	0.7	0.7	0.8	0.9	0.9	0.9	1.0	1.1	1.2	1.2	1.3	1.4	1.4	1.5	1.5	1.6	1.6	1.6	1.7	1.7	1.7
	VI	1235	σ	0.2	0.3	0.3	0.4	0.4	0.5	0.5	0.6	0.6	0.6	0.7	0.7	0.8	0.8	0.8	0.9	0.9	0.9	1.0	1.0	1.0	1.1	1.1	1.1	1.1
	I–VI	6252	σ	0.1	0.1	0.2	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.4	0.4	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.6	0.6

TABLE II

Average Number of Units in Bases Used in Computing Percentages of Timing Tendency for Each Observer Compared with Three Other Observers in Series A and Standard Deviations of Percentages of These Bases, by Film and Category

Category	Film	Average Number of Units in Base by Film	%	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	6.0	7.0	8.0	9.0	10.0	11.0	12.0	13.0	14.0	15.0	16.0	17.0	18.0	19.0	20.0
				Standard Deviations of Each of the Following Percentages on the Corresponding Base																								
Material	I	1316	σ	0.2	0.3	0.3	0.4	0.4	0.5	0.5	0.6	0.6	0.6	0.7	0.7	0.8	0.8	0.8	0.9	0.9	0.9	1.0	1.0	1.0	1.1	1.1	1.1	1.1
	II	1771	σ	0.2	0.2	0.3	0.3	0.4	0.4	0.4	0.5	0.5	0.5	0.6	0.6	0.7	0.7	0.7	0.8	0.8	0.8	0.8	0.9	0.9	0.9	0.9	1.0	1.0
	III	665	σ	0.3	0.4	0.5	0.5	0.6	0.7	0.7	0.8	0.8	0.8	0.9	1.0	1.0	1.1	1.2	1.2	1.3	1.3	1.3	1.4	1.4	1.5	1.5	1.5	1.5
	IV	1590	σ	0.2	0.2	0.3	0.3	0.4	0.4	0.4	0.5	0.5	0.5	0.6	0.6	0.7	0.7	0.7	0.8	0.8	0.8	0.8	0.9	0.9	0.9	0.9	1.0	1.0
	V	1978	σ	0.1	0.2	0.3	0.3	0.3	0.4	0.4	0.4	0.5	0.5	0.5	0.6	0.6	0.6	0.7	0.7	0.7	0.8	0.8	0.8	0.8	0.8	0.9	0.9	0.9
	VI	2772	σ	0.1	0.2	0.2	0.3	0.3	0.3	0.4	0.4	0.4	0.4	0.5	0.5	0.5	0.6	0.6	0.6	0.6	0.7	0.7	0.7	0.7	0.7	0.8	0.8	0.8
	I–VI	10092	σ	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.4	0.4	0.4
Self	I	8893	σ	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.4	0.4	0.4
	II	14871	σ	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
	III	10703	σ	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.4
	IV	17540	σ	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
	V	12778	σ	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.4
	VI	14817	σ	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
	I–VI	79602	σ	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Person	I	2368	σ	0.1	0.2	0.2	0.3	0.3	0.3	0.4	0.4	0.4	0.4	0.5	0.5	0.5	0.6	0.6	0.6	0.6	0.7	0.7	0.7	0.7	0.7	0.8	0.8	0.8
	II	5198	σ	0.1	0.1	0.2	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.4	0.4	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.6	0.6
	III	1667	σ	0.2	0.2	0.3	0.3	0.4	0.4	0.4	0.5	0.5	0.5	0.6	0.6	0.7	0.7	0.7	0.8	0.8	0.8	0.8	0.9	0.9	0.9	0.9	1.0	1.0
	IV	4278	σ	0.1	0.1	0.2	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.4	0.4	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.6	0.6
	V	1541	σ	0.2	0.2	0.3	0.3	0.4	0.4	0.4	0.5	0.5	0.5	0.6	0.6	0.7	0.7	0.7	0.8	0.8	0.8	0.8	0.9	0.9	0.9	0.9	1.0	1.0
	VI	3704	σ	0.1	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.4	0.5	0.5	0.5	0.5	0.6	0.6	0.6	0.6	0.6	0.6	0.7	0.7	0.7
	I–VI	18756	σ	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.3

[illegible] 4

Princeton Theological Seminary Library
1 1012 01600 7116